



# **STIC Search Report**

## **Biotech-Chem Library**

**STIC Database Tracking Number: 138057**

**TO: James Schultz**  
**Location: rem/2d18/2c18**  
**Art Unit: 1635**  
**Monday, November 22, 2004**

**Case Serial Number: 10/001851**

**From: Alex Waclawiw**  
**Location: Biotech-Chem Library**  
**Rem 1A71**  
**Phone: 272-2534**

**[Alexandra.waclawiw@uspto.gov](mailto:Alexandra.waclawiw@uspto.gov)**

### **Search Notes**

This Page Blank (uspto)

STIC-Biotech/ChemLib

138057

mej

From: Schultz, James  
Sent: Wednesday, November 17, 2004 8:52 AM  
To: STIC-Biotech/ChemLib  
Subject: Seq Search 10/001,851

Hello,  
Please run a standard amino acid search on SEQ ID NO: 2 from the above entitled application. Instead of truncating the results at the top 15, could you please return the top 50 results from each database as well?  
Thanks,  
Doug Schultz

James Douglas Schultz, PhD  
AU 1635 (Biotechnology)  
Patent Examiner  
United States Patent and Trademark Office  
(Office) REM 2D18  
(Mail) REM 2C18  
(571) 272-0763

RECEIVED  
NOV 17 2004  
STIC-BIOTECH/ChemLib  
(STIC)

Point of Contact:

Alexandra Waclawiw

Technical Info. Specialist

\*\*\*\*\*  
STAFF USE ONLY

~~GM1-6A02 Tel: 308-449~~

Searcher: \_\_\_\_\_  
Searcher Phone: 2-\_\_\_\_\_  
Date Searcher Picked up: 11-22  
Date Completed: \_\_\_\_\_  
Searcher Prep/Rev. Time: 11-22  
Online Time: \_\_\_\_\_

\*\*\*\*\*

Type of Search

NA Sequence: # \_\_\_\_\_  
AA Sequence: # (1)  
Structure: # \_\_\_\_\_  
Bibliographic: \_\_\_\_\_  
Litigation: \_\_\_\_\_  
Patent Family: \_\_\_\_\_  
Other: \_\_\_\_\_

\*\*\*\*\*

Vendors and cost where applicable

STN: \_\_\_\_\_  
DIALOG: \_\_\_\_\_  
QUESTEL/ORBIT: \_\_\_\_\_  
LEXIS/NEXIS: \_\_\_\_\_  
SEQUENCE SYSTEM: \_\_\_\_\_  
WWW/Internet: \_\_\_\_\_  
Other(Specify): \_\_\_\_\_

**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

## OM protein - protein search, using sw model

Run on: November 22, 2004, 13:36:41 ; Search time 68 Seconds

(without alignments)  
3181.085 Million cell updates/sec

Title: US-10-001-851-2

Perfect score: 3278

Sequence: 1 MRKKEKLLQAVLVLAALV.....TQQWLFPHNTSYLAEKRN 603

Scoring table: BIOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2002273 seqs, 358729239 residues

Total number of hits satisfying chosen parameters: 2002273

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database :

1: A\_GeneSeq\_23Sep04:\*  
2: geneSeqp1980s:\*  
3: geneSeqp1990s:\*  
4: geneSeqp2000s:\*  
5: geneSeqp2001s:\*  
6: geneSeqp2002s:\*  
7: geneSeqp2003as:\*  
8: geneSeqp2003bs:\*  
9: geneSeqp2004s:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query	Match	Length	ID	Description
1	3278	100.0	603	4	AAU07778	AAU07778 Human nov
2	3278	100.0	603	5	AAU99098	AAU99098 Human gly
3	3272	99.8	603	7	ADG35387	ADG35387 Human N-a
4	3242.5	98.9	631	4	AAU07777	AAU07777 Human nov
5	2771	84.5	506	6	AAU07772	AAU07772 Human nov
6	2771	84.5	506	6	AAU07781	AAU07781 Carboxydr
7	2746.5	83.8	535	4	AAU07771	AAU07771 Human nov
8	1893	57.7	339	4	AAU94733	AAU94733 Human pro
9	1877	57.3	366	4	AAU07776	AAU07776 Human nov
10	1818	55.5	407	4	AAU07776	AAU07776 Human nov
11	1403.5	42.8	319	4	ABG08187	ABG08187 Human ace
12	1396	42.6	275	7	ADG25865	ADG25865 Human N-a
13	1392	42.5	275	8	ADG17613	ADG17613 Human nov
14	1385.5	42.3	335	4	ABG08189	ABG08189 Human nov
15	1370	41.8	269	4	AAU07770	AAU07770 Human nov
16	1360	41.5	666	4	ABG08187	ABG08187 Human nov
17	1360	41.5	666	4	ABG08187	ABG08187 Human nov
18	1311	40.0	321	4	AAU07774	AAU07774 Human nov
19	1186	36.2	1252	4	ABG08190	ABG08190 Human nov
20	1155	35.2	499	7	ADG65304	ADG65304 Human pol
21	1125	34.3	559	7	ADG65304	ADG65304 Human pol
22	1125	34.3	559	7	ADG65304	ADG65304 Human pol
23	1116	34.0	559	7	ADG65304	ADG65304 Human pol
24	1116	34.0	559	7	ADG65304	ADG65304 Human pol
25	1116	34.0	559	7	ADG65304	ADG65304 Human pol

26	1115	34.0	517	2	AA66402	AA66402 GalNac-tr
27	1115	34.0	517	2	AA66402	AA66402 Honeybee
28	1115	34.0	517	2	AA66402	AA66402 Cattle Ga
29	1115	34.0	517	2	AA66402	AA66402 GalNac-tr
30	1115	34.0	517	2	AA66402	AA66402 Bovine N-
31	1113.5	34.0	630	4	ABG64271	ABG64271 Drosophi
32	1112	33.9	571	7	ADG10116	ADG10116 Human NOV
33	1106	33.7	556	5	ABG32379	ABG32379 Human NOV
34	1106	33.7	556	5	ABG32379	ABG32379 Human NOV
35	1106	33.7	556	5	ABG32379	ABG32379 Human NOV
36	1096.5	33.5	555	7	ADG10114	ADG10114 Human NOV
37	1085.5	33.1	573	8	ADG04927	ADG04927 Sea squi
38	1050	32.0	657	4	AA66402	AA66402 Human mem
39	1048	32.0	478	6	ABR41518	ABR41518 Human DIT
40	1048	32.0	478	6	ADL22696	ADL22696 Human dis
41	1045.5	31.9	658	4	AAU78711	AAU78711 Human pro
42	1040	31.7	930	7	ADG55070	ADG55070 Rat Prote
43	1035.5	31.6	591	4	ABG62319	ABG62319 Drosophi
44	1026	31.3	578	6	ABU07527	ABU07527 Human N-a
45	1024.5	31.3	188	7	ADG25974	ADG25974 N-acetyl
46	1015	31.0	610	7	ADG55072	ADG55072 Human pro
47	1013	30.9	940	3	AAU53040	AAU53040 Human sec
48	1013	30.9	940	8	ADK70539	ADK70539 Respirato
49	999.5	30.5	558	5	ABP69229	ABP69229 Human pol
50	999.5	30.5	558	5	AAE25021	AAE25021 Human dr
51	991.5	30.2	581	5	ABG32510	ABG32510 Human gly
52	991.5	30.2	581	5	AAE25019	AAE25019 Human dr
53	991.5	30.2	581	5	ADH48838	ADH48838 NOV52 pro
54	990	30.2	240	4	AAU07775	AAU07775 Human nov
55	988	30.1	209	4	AAU07775	AAU07775 Human nov
56	986.5	30.1	581	7	ADG35388	ADG35388 Human N-a
57	986.5	30.1	581	8	ADG20087	ADG20087 Human pro
58	977.5	29.8	560	7	ADG25922	ADG25922 Novel N-a
59	971.5	29.6	581	7	ADG25914	ADG25914 Novel N-a
60	971.5	29.6	552	4	AAE06620	AAE06620 Human pro
61	971	29.6	552	4	AAU12269	AAU12269 Human pro
62	971	29.6	552	4	AAU09026	AAU09026 Human nov
63	971	29.6	552	6	ABO17713	ABO17713 Novel hum
64	971	29.6	552	6	ABR58585	ABR58585 Human can
65	971	29.6	552	6	ABU08097	ABU08097 Human pro
66	971	29.6	552	6	ABU66667	ABU66667 Human pro
67	971	29.6	552	6	ABU59748	ABU59748 Novel sec
68	971	29.6	552	6	ABO24938	ABO24938 Human sec
69	971	29.6	552	6	ABU66943	ABU66943 Human sec
70	971	29.6	552	6	ADG45715	ADG45715 Novel hum
71	971	29.6	552	6	ADG476146	ADG476146 Human pro
72	971	29.6	552	6	ADG18796	ADG18796 Human pro
73	971	29.6	552	6	ADG61419	ADG61419 Homo sapi
74	971	29.6	552	6	ADG19204	ADG19204 Novel hum
75	971	29.6	552	6	ADG27745	ADG27745 Human pro
76	971	29.6	552	6	ADG66224	ADG66224 Novel hum
77	971	29.6	552	6	ADG15788	ADG15788 Human pro
78	971	29.6	552	6	ADG47574	ADG47574 Human pro
79	971	29.6	552	6	ADG67369	ADG67369 Human pro
80	971	29.6	552	6	ADG30376	ADG30376 Human pro
81	971	29.6	552	6	ADG85672	ADG85672 Novel hum
82	971	29.6	552	6	ADG85684	ADG85684 Human pro
83	971	29.6	552	6	ADG79188	ADG79188 Human pro
84	971	29.6	552	6	ADG71327	ADG71327 Novel hum
85	971	29.6	552	6	ADG16529	ADG16529 Human pro
86	971	29.6	552	6	ADG161621	ADG161621 Novel hum
87	971	29.6	552	6	ADG14684	ADG14684 Human pro
88	971	29.6	552	6	ADG18645	ADG18645 Novel hum
89	971	29.6	552	6	ADG193860	ADG193860 Human pro
90	971	29.6	552	6	ADG19756	ADG19756 Novel hum
91	971	29.6	552	6	ADG13068	ADG13068 Human pro
92	971	29.6	552	6	ADG43246	ADG43246 Novel hum
93	971	29.6	552	6	ADG47322	ADG47322 Human pro
94	971	29.6	552	6	ADG24555	ADG24555 Human pro
95	971	29.6	552	6	ADG28079	ADG28079 Human pro
96	971	29.6	552	6	ADG75042	ADG75042 Human pro
97	971	29.6	552	6	ADG85120	ADG85120 Novel hum
98	971	29.6	552	6	ADG84568	ADG84568 Novel hum

99 971 29.6 552 6 ADB29824 Human PRO  
100 971 29.6 552 6 ADA80352 Human PRO

## ALIGNMENTS

RESULT 1  
AAU07778  
ID AAU07778 standard; protein; 603 AA.

AC AAU07778;

DT 04-DEC-2001 (first entry)

DE Human novel transferase protein, NHP #21.

KW Human; transferase; breast cancer; prostate cancer; immunogen;

KM gene therapy; antisense.

OS Homo sapiens.

PN WC020164903-A2.

PD 07-SEP-2001.

PF 28-FEB-2001; 2001WO-US006460.

PR 29-FEB-2000; 2000US-0185920P.

PR 02-MAR-2000; 2000US-0186558P.

PR 24-MAR-2000; 2000US-0191849P.

XX (LEXI-) LEXICON GENETICS INC.

PI Donoho G, Hilbun E, Turner CA, Friedrich G, Abuin A;

PI Zambrowicz B, Sands AT, Walke DW, Wilganowski NL, Hu Y, Kleke JA;

PI Potter DG;

DR WPI: 2001-550185/61.

DR N-PSDB; AAS12625.

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

Sequence 603 AA;

Query Match 100.0%; Score 3278; DB 4; Length 603;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 603; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MRRKREKLLQVALVLAALVLLPNTGLMALYREROPDGTGGSGAAVAPAAQSGSHSROK 60  
DB 1 MRRKREKLLQVALVLAALVLLPNTGLMALYREROPDGTGGSGAAVAPAAQSGSHSROK 60  
QY 61 KTFPLGDCQKLDMDKDKAIRDQAVNGGEGRPYMTDARVVOAARENGFNLYSDK 120  
DB 61 KTFPLGDCQKLDMDKDKAIRDQAVNGGEGRPYMTDARVVOAARENGFNLYSDK 120  
QY 121 ISLNRSLPDIRHPNCNKRYLETLPTNTSIIIPFHNMGSSLLRTVHSLNRSPELVAEI 180

DB 121 ISLNRSLPDIRHPNCNKRYLETLPTNTSIIIPFHNMGSSLLRTVHSLNRSPELVAEI 180

QY 181 VLVDPSGDRHLKCPLEDYALFPSVRILRPKKEGLIRPMLGASVATGVITFLSHC 240

DB 181 VLVDPSGDRHLKCPLEDYALFPSVRILRPKKEGLIRPMLGASVATGVITFLSHC 240

QY 241 EANYNMLPPLLDRIARRKTIIVCPMIDVIDHDDPRYETQAGDAMRGAFDMEYKRLIP 300

DB 241 EANYNMLPPLLDRIARRKTIIVCPMIDVIDHDDPRYETQAGDAMRGAFDMEYKRLIP 300

QY 301 PELQKADPSDFESPVNAGLFAVDRKFWELGYPDGLIWSGEQYEISFKVMCGGRM 360

DB 301 PELQKADPSDFESPVNAGLFAVDRKFWELGYPDGLIWSGEQYEISFKVMCGGRM 360

QY 361 EDIPCSRVGHTYRKYVYKYPAGVSLARNLKRVAEVMWDEYAEIYQRRPEYRHLSDGV 420

DB 361 EDIPCSRVGHTYRKYVYKYPAGVSLARNLKRVAEVMWDEYAEIYQRRPEYRHLSDGV 420

QY 421 AVOKKLRSSLNCKSFKFMFTKIAMDLPKFYPPVPPAAAMGEIRNVTGLCADTKHGALG 480

DB 421 AVOKKLRSSLNCKSFKFMFTKIAMDLPKFYPPVPPAAAMGEIRNVTGLCADTKHGALG 480

QY 481 SPLRLGCGVCRGSAAMNNQVFTFTWRBDIRPDPOHTKKCFDAISHTSPVTLYDCHS 540

DB 481 SPLRLGCGVCRGSAAMNNQVFTFTWRBDIRPDPOHTKKCFDAISHTSPVTLYDCHS 540

QY 541 MKGNQMKYRKDKTLHPVSGSCMDSCSDHRIIPMTNCPBSLTOQMLFEHTNSTVLEKF 600

DB 541 MKGNQMKYRKDKTLHPVSGSCMDSCSDHRIIPMTNCPBSLTOQMLFEHTNSTVLEKF 600

QY 601 NRN 603

DB 601 NRN 603

RESULT 2

AAU99098 standard; protein; 603 AA.

AAU99098;

DT 30-AUG-2002 (first entry)

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

Human glycosyl transferase 47169.

Human; glycosyl transferase; 47169; 33935; cancer; carbohydrate storage;

diabetes mellitus; hypoglycaemia; arthritis; rheumatism;

autoimmune disorder; systemic lupus erythematosus; Grave's disease;

myaesthesia gravis; insulin resistance; scleroderma; rheumatoid arthritis;

autoimmune infertility; tumorigenesis.

OS Homo sapiens.

PN WC0200240657-A2.

PD 23-MAY-2002.

PF 20-NOV-2001; 2001WO-US047575.

PR 20-NOV-2000; 2000US-0249939P.

PR (MILL-) MILLENNIUM PHARM INC.

PI Meyers R, Williamson M;

PI WPI: 2002-508326/54.

PI N-PSDB; ABR86093.

Use of modulator of activity of novel glycosyl transferase proteins, 47169/33935 proteins, for making a medicament for modulating ability of cell to affect glycosylation state of lipid or polypeptide target in cell.

XX  
PS Claim 27; Fig 1; 153pp; English.  
CC The invention relates to the use of a modulator of the activity of an  
CC enzyme such as 47169 or 33935 protein (a novel glycosyl transferase  
CC protein) for making a medicament for modulating the ability of a cell to  
CC affect the glycosylation state of a lipid target or polypeptide target in  
CC a cell. Also included is a method of assessing (M1) if a test compound is  
CC useful for modulating at least one phenomenon (P) such as non-covalent  
CC binding between a protein and one of a cell, a virus and another protein;  
CC cell signaling, cell differentiation, tumorigenesis, cell adhesion, cell  
CC motility, cell-to-cell interaction, cell invasivity, cell proliferation,  
CC gene transcription, and an immune response, comprising: (a) adding the  
CC test compound to a first composition comprising a 603 residue 47169  
CC polypeptide sequence (S2), or a sequence at least 90 % identical to a 492  
CC residue 33935 polypeptide sequence (S12), both given in the specification  
CC ; and (b) comparing the activity in the first composition and in a second  
CC composition that is substantially identical to the first composition,  
CC except that it lacks the test compound, whereby a difference in the  
CC activity in the first and second compositions is an indication that the  
CC test compound is useful for modulating the phenomenon. The method is  
CC useful for making a medicament for modulating the ability of a cell (e.g.  
CC human endothelial cell such as lung cell, breast cell or colon cell,  
CC preferably a tumour cell) to affect the glycosylation state of a target  
CC such as lipid or polypeptide. The identified test compound is useful for  
CC treating disorders such as diabetes mellitus, hypoglycaemia, arthritis,  
CC rheumatism, autoimmune disorders (e.g. systemic lupus erythematosus,  
CC Grave's disease, myasthenia gravis, insulin resistance, rheumatoid  
CC arthritis, scleroderma and autoimmune infertility), tumorigenesis,  
CC cancer and tumour metastasis. The present sequence represents glycosyl  
CC transferase 47169  
CC  
SQ Sequence 603 AA;  
Query Match 100.0%; Score 3278; DB 5; Length 603;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 603; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MRRERKRLQAVLVLAALVTLVNVGIMALYRERQPDGPGSGAVALPAAGGSHSRQK 60  
DB 1 MRRERKRLQAVLVLAALVTLVNVGIMALYRERQPDGPGSGAVALPAAGGSHSRQK 60  
QY 61 KTFPLDGGQKLDKMDHDEALRDQVRNGEQRPRYMTDAERVDQYRENGFRIYVSDK 120  
DB 61 KTFPLDGGQKLDKMDHDEALRDQVRNGEQRPRYMTDAERVDQYRENGFRIYVSDK 120  
QY 121 ISLNRSLPDIRHPNCKRKYLETLPNTSIIIPFNEGSSILRTVHSLNRPPELVAEI 180  
DB 121 ISLNRSLPDIRHPNCKRKYLETLPNTSIIIPFNEGSSILRTVHSLNRPPELVAEI 180  
QY 121 ISLNRSLPDIRHPNCKRKYLETLPNTSIIIPFNEGSSILRTVHSLNRPPELVAEI 180  
DB 121 ISLNRSLPDIRHPNCKRKYLETLPNTSIIIPFNEGSSILRTVHSLNRPPELVAEI 180  
QY 181 VLVDPSDRHLKPLEDYMALPPSVRIILTKKREGILRTMIGASVATGDIITFLDSHC 240  
DB 181 VLVDPSDRHLKPLEDYMALPPSVRIILTKKREGILRTMIGASVATGDIITFLDSHC 240  
QY 181 VLVDPSDRHLKPLEDYMALPPSVRIILTKKREGILRTMIGASVATGDIITFLDSHC 240  
DB 181 VLVDPSDRHLKPLEDYMALPPSVRIILTKKREGILRTMIGASVATGDIITFLDSHC 240  
QY 241 EAVNNMPLPILDIRAARKTIVCMIDVIDDHFRTYQADNARGAFDMWYTKRPIR 300  
DB 241 EAVNNMPLPILDIRAARKTIVCMIDVIDDHFRTYQADNARGAFDMWYTKRPIR 300  
QY 241 EAVNNMPLPILDIRAARKTIVCMIDVIDDHFRTYQADNARGAFDMWYTKRPIR 300  
DB 241 EAVNNMPLPILDIRAARKTIVCMIDVIDDHFRTYQADNARGAFDMWYTKRPIR 300  
QY 301 PELQKADPDPFESPVWAGGLFAVDRKFMELGSDYDGLTIGGEQVEISPKVMCCGRM 360  
DB 301 PELQKADPDPFESPVWAGGLFAVDRKFMELGSDYDGLTIGGEQVEISPKVMCCGRM 360  
QY 301 PELQKADPDPFESPVWAGGLFAVDRKFMELGSDYDGLTIGGEQVEISPKVMCCGRM 360  
DB 301 PELQKADPDPFESPVWAGGLFAVDRKFMELGSDYDGLTIGGEQVEISPKVMCCGRM 360  
QY 361 EDIPCSRVGHLYRKYVPYKPVAGVSLANLKRVAEVMDEYAEYIYORRPEYRLHLSADV 420  
DB 361 EDIPCSRVGHLYRKYVPYKPVAGVSLANLKRVAEVMDEYAEYIYORRPEYRLHLSADV 420  
QY 361 EDIPCSRVGHLYRKYVPYKPVAGVSLANLKRVAEVMDEYAEYIYORRPEYRLHLSADV 420  
DB 361 EDIPCSRVGHLYRKYVPYKPVAGVSLANLKRVAEVMDEYAEYIYORRPEYRLHLSADV 420  
QY 421 AVOKKSLSSLNCKSFKKFMKTIAMDLPKFPVPEPPAAAMEIINNGVCADTRKHALG 480  
DB 421 AVOKKSLSSLNCKSFKKFMKTIAMDLPKFPVPEPPAAAMEIINNGVCADTRKHALG 480  
QY 421 AVOKKSLSSLNCKSFKKFMKTIAMDLPKFPVPEPPAAAMEIINNGVCADTRKHALG 480  
DB 421 AVOKKSLSSLNCKSFKKFMKTIAMDLPKFPVPEPPAAAMEIINNGVCADTRKHALG 480  
QY 481 SPLRLGECVGRGSAANNMNVQVFTFTRREDIRPDGPHTTKFCFDAQSHSTPTLYDCHS 540  
DB 481 SPLRLGECVGRGSAANNMNVQVFTFTRREDIRPDGPHTTKFCFDAQSHSTPTLYDCHS 540

QY 541 MKGNQMKRKDKXTLYHPVSGCWDCESDHRIEMNTCNPSLTOQWLFHTNSTYLEKF 600  
DB 541 MKGNQMKRKDKXTLYHPVSGCWDCESDHRIEMNTCNPSLTOQWLFHTNSTYLEKF 600  
QY 601 NRN 603  
DB 601 NRN 603  
RESULT 3  
ID ADC35387 standard; protein; 603 AA.  
XX  
AC ADC35387;  
XX  
DT 18-DEC-2003 (first entry)  
XX  
DB Human N-acetylglucosamine transferase (GalNAc) T13 protein.  
XX  
KW N-acetylglucosamine transferase; GalNAc; alpha1-bond;  
KW genetic engineering; GalNAc-T13; cytosolic; cancer treatment;  
KW cancer diagnosis; gene therapy; human; enzyme.  
XX  
OS Homo sapiens.  
XX  
PN WO2003057897-A1.  
XX  
PD 17-JUL-2003.  
XX  
PF 06-JAN-2003; 2003WO-JP000008.  
XX  
PR 28-DEC-2001; 2001JP-00401507.  
PR 05-JUN-2002; 2002JP-00163832.  
PR 12-JUL-2002; 2002JP-00203696.  
XX  
PA (NAAD-) NAT INST ADVANCED IND SCI & TECHNOLOGY.  
PA (AMSH) AMERSHAM BIOSCIENCES KK.  
PA (FORE) FUJIREBIO INC.  
XX  
PI Narimatsu H, Zhang Y, Gotoh M;  
XX  
DR WPI; 2003-587133/55.  
DR N-PSDB; ADC35394.  
XX  
PT Novel UDP-N-acetyl-D-galactosamine:polypeptide N-acetylglucosamine  
PT transferase and encoded nucleic acid, applicable in identifying O-bonded  
PT sugar-attached proteins, gene diagnosis, and cancer therapy.  
XX  
PS Claim 1; SEQ ID NO 3; 153pp; Japanese.  
XX  
XX This invention relates to a novel protein with an activity of  
CC transferring N-acetylglucosamine (GalNAc) via an alpha1-bond to a  
CC hydroxyl group of serine or threonine in a protein or a peptide sequence.  
CC The invention provides gene and protein sequences for these enzymes to  
CC enable genetic engineering or production of these enzymes. Seven genes  
CC and their proteins are described by the invention, GalNAc-T11 to 17. The  
CC enzymes of the invention may have cytosolic activity, and hence may be  
CC used for treatment or diagnosis of cancer, and the sequences described  
CC may be useful in gene therapy. The present sequence is the partial  
CC sequence of the human GalNAc-T13 transferase protein of the invention.  
XX  
SQ Sequence 603 AA;  
Query Match 99.8%; Score 3272; DB 7; Length 603;  
Best Local Similarity 99.8%; Pred. No. 0;  
Matches 602; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MRRERKRLQAVLVLAALVTLVNVGIMALYRERQPDGPGSGAVALPAAGGSHSRQK 60  
DB 1 MRRERKRLQAVLVLAALVTLVNVGIMALYRERQPDGPGSGAVALPAAGGSHSRQK 60  
QY 61 KTFPLDGGQKLDKMDHDEALRDQVRNGEQRPRYMTDAERVDQYRENGFRIYVSDK 120

Db 61 KTFPLGDCQKLDHMDKEAIRDAQRVNGEGRPYMTDAERVDQAVRENGFNIVYSDK 120  
Qy 121 ISLRSLEPDIRHPNCSKRYLETLPNTSIIIPFNHEGSSILRTVHSLNRPPELVAEI 180  
Db 121 ISLRSLEPDIRHPNCSKRYLETLPNTSIIIPFNHEGSSILRTVHSLNRPPELVAEI 180  
Qy 181 VLVDPSDRHLKKPLBEDYMLFPSVRLRTKKEGILRTMTLGASVATGVITFLSDHC 240  
Db 181 VLVDPSDRHLKKPLBEDYMLFPSVRLRTKKEGILRTMTLGASVATGVITFLSDHC 240  
Qy 241 EANNVMLPPLIDRLIARNKTIIVCPMIDVIDHDDFRYETQADANRGAFFDMETTKRIP 300  
Db 241 EANNVMLPPLIDRLIARNKTIIVCPMIDVIDHDDFRYETQADANRGAFFDMETTKRIP 300  
Qy 301 PELQKADSPDPESPVNAGGLFAVDRKMFWEIGYDPLGLTWGGEQYEISFKVMCGGRM 360  
Db 301 PELQKADSPDPESPVNAGGLFAVDRKMFWEIGYDPLGLTWGGEQYEISFKVMCGGRM 360  
Qy 361 EDICSRVGHITRYKVPYKVPAGVSLAANLKRVAEVMDEYAEIYQRRPEYRHLASDV 420  
Db 361 EDICSRVGHITRYKVPYKVPAGVSLAANLKRVAEVMDEYAEIYQRRPEYRHLASDV 420  
Qy 421 AVOKKLRSSLNCKSFMTKIAMDLPRFYPPVEPPAAAGBIRNVGTGLCADTYHGALG 480  
Db 421 AVOKKLRSSLNCKSFMTKIAMDLPRFYPPVEPPAAAGBIRNVGTGLCADTYHGALG 480  
Qy 481 SPLLEGCVRGGEAANNMVOFTFTWRBDIRPGDPQHTKKRCPAISHTSPTVLTYDCHS 540  
Db 481 SPLLEGCVRGGEAANNMVOFTFTWRBDIRPGDPQHTKKRCPAISHTSPTVLTYDCHS 540  
Qy 541 MKGNQLMWKRYKDKTLHYFVSGSCMDSCSDHRIFNATCNPSLSLTQOMLFHTNSTVLEKF 600  
Db 541 MKGNQLMWKRYKDKTLHYFVSGSCMDSCSDHRIFNATCNPSLSLTQOMLFHTNSTVLEKF 600  
Qy 601 NRN 603  
Db 601 NRN 603

RESULT 4  
AAU07777  
ID AAU07777 standard; protein; 631 AA.  
XX  
AC AAU07777;  
XX  
DT 04-DEC-2001 (first entry)  
XX  
DE Human novel transferase protein, NHP #20.  
XX  
KW Human; transferase; breast cancer; prostate cancer; immunogen;  
XX  
KW gene therapy; antisense.  
XX  
OS Homo sapiens.  
XX  
FH Key Location/Qualifiers  
FT Misc-difference 630..631  
FT /note= "Encoded by AATGGAAC"  
XX  
XX WO200164903-A2.  
XX  
PD 07-SEP-2001.  
XX  
PF 28-FEB-2001; 2001MO-US006460.  
XX  
PR 29-FEB-2000; 2000US-0185920P.  
PR 02-MAR-2000; 2000US-0186558P.  
PR 24-MAR-2000; 2000US-0191849P.  
XX  
PA (LEXI-) LEXICON GENETICS INC.  
XX  
PI Donoho G, Hilbun E, Turner CA, Friedrich G, Abuin A,  
Zambrowicz B, Sands AT, Walke DW, Wilganowski NL, Hu Y, Kleke JA,

PI Potter DG;  
XX WPI; 2001-550185/61.  
DR N-P6DB; AAS12624.  
XX  
XX Novel nucleic acid sequences encoding novel human proteins useful for  
PT diagnosis, drug screening, clinical trial monitoring and treatment of  
PT diseases and disorders.  
XX  
PS Claim 6; Page 54-55; 60pp; English.  
XX  
CC The invention relates to isolated nucleic acids encoding novel human  
CC transferase proteins (NHP). The nucleic acids and proteins are useful for  
CC diagnosis, drug screening, clinical trial monitoring and treatment of  
CC diseases and disorders e.g. breast and prostate cancer. NHPs can also be  
CC useful for augmenting the efficacy of chemotherapeutic agents used in  
CC treatment of breast or prostate cancer. The nucleic acid is also useful  
CC in NHP gene regulation, and as antisense primers in amplification  
CC reactions of NHP gene sequences. NHPs are useful for producing  
CC antibodies. The present sequence represents a novel human transferase  
CC which has sequence similarity to N-acetyl-galactosaminyltransferase  
SQ Sequence 631 AA;  
Query Match 98.9%; Score 3242.5; DB 4; Length 631;  
Best Local Similarity 95.4%; Pred. No. 0;  
Matches 601; Conservative 0; Mismatches 0; Indels 29; Gaps 1;  
Qy 1 MRKREKLLQVALVLAALVTLIPNVGIMALYREOPDPTGGSGAAVAPAGGSHSRQK 60  
Db 1 MRKREKLLQVALVLAALVTLIPNVGIMALYREOPDPTGGSGAAVAPAGGSHSRQK 60  
Qy 61 KTFPLGDCQKLDHMDKEAIRDAQRVNGEGRPYMTDAERVDQAVRENGFNIVYSDK 120  
Db 61 KTFPLGDCQKLDHMDKEAIRDAQRVNGEGRPYMTDAERVDQAVRENGFNIVYSDK 120  
Qy 121 ISLRSLEPDIRHPNCSKRYLETLPNTSIIIPFNHEGSSILRTVHSLNRPPELVAEI 180  
Db 121 ISLRSLEPDIRHPNCSKRYLETLPNTSIIIPFNHEGSSILRTVHSLNRPPELVAEI 180  
Qy 181 VLVDPSDRHLKKPLBEDYMLFPSVRLRTKKEGILRTMTLGASVATGVITFLSDHC 240  
Db 181 VLVDPSDRHLKKPLBEDYMLFPSVRLRTKKEGILRTMTLGASVATGVITFLSDHC 240  
Qy 241 EANNVMLPPLIDRLIARNKTIIVCPMIDVIDHDDFRYETQADANRGAFFDMETTKRIP 300  
Db 241 EANNVMLPPLIDRLIARNKTIIVCPMIDVIDHDDFRYETQADANRGAFFDMETTKRIP 300  
Qy 301 PELQKADSPDPESPVNAGGLFAVDRKMFWEIGYDPLGLTWGGEQYEISFK----- 352  
Db 301 PELQKADSPDPESPVNAGGLFAVDRKMFWEIGYDPLGLTWGGEQYEISFK----- 352  
Qy 353 -----VNMCGRMEDIPCSRGHITRYKVPYKVPAGVSLAANL 391  
Db 353 -----VNMCGRMEDIPCSRGHITRYKVPYKVPAGVSLAANL 391  
Qy 361 VSNMPOAVFLPRAIPNMLALQVMWCGRMEDIPCSRGHITRYKVPYKVPAGVSLAANL 420  
Db 361 VSNMPOAVFLPRAIPNMLALQVMWCGRMEDIPCSRGHITRYKVPYKVPAGVSLAANL 420  
Qy 392 RVAEVMDEYAEIYQRRPEYRHLASDVAVOKKLRSSLNCKSFMTKIAMDLPRFY 451  
Db 392 RVAEVMDEYAEIYQRRPEYRHLASDVAVOKKLRSSLNCKSFMTKIAMDLPRFY 451  
Qy 421 RVAEVMDEYAEIYQRRPEYRHLASDVAVOKKLRSSLNCKSFMTKIAMDLPRFY 480  
Db 421 RVAEVMDEYAEIYQRRPEYRHLASDVAVOKKLRSSLNCKSFMTKIAMDLPRFY 480  
Qy 452 PVBPAAAGBIRNVGTGLCADTYHGALGSPRLLEGCVRGGEAANNMVOFTFTWRDI 511  
Db 452 PVBPAAAGBIRNVGTGLCADTYHGALGSPRLLEGCVRGGEAANNMVOFTFTWRDI 511  
Qy 512 RPDGPHTKKRCFPAISHTSPTVLTYDCHSMKGNQLMWKRYKDKTLHYFVSGSCMDSCSDH 571  
Db 512 RPDGPHTKKRCFPAISHTSPTVLTYDCHSMKGNQLMWKRYKDKTLHYFVSGSCMDSCSDH 571  
Qy 572 RIFMNTCNPSLSLTQOMLFHTNSTVLEKEN 601  
Db 572 RIFMNTCNPSLSLTQOMLFHTNSTVLEKEN 601

ID	Accession	Description
Oy	338	GLEIIMGGEQVSEISPEFVMNMGGMEDIPCSRVIHITRKVPYKPVPGVSLABNLKRVAEW 397
Dd	241	GLEIHWGSGQVYEISPEFVMNMGGMEDI PCSRVHIITRKVPYKPVPGVSLARNLKRAEWA 3000
Oy	398	MDEVALEYIQRRPPEPRHLASAGVAQAOKKLRSLLNCKSPFKPMFTKILAMDLPKEFPVEPPA 457
Dd	301	MDEVALEYIQRRPPEPRHLASAGDVAVQKKLRSSLNCKSPFKPMFTKILAMDLPKEFPVEPPA 3666
Oy	458	AAMGEIRNVGTGLCADTGHGALIGSPLRLIEGCVRGEGEAANNMQUPTFTWRBIDIRPDQP 517
Dd	361	AAMGEIRNVGTGLCADTGHGALIGSPLRLIEGCVRGEGEAANNMQUPTFTWRBIDIRPDQP 4200
Oy	518	HTKKRCPFAIHSTHSVTLYDCHSMKGNOLMKYRKDKTLYPHPSSGSCMDSESHPRIEMNT 5777
Dd	421	HTKKRCPFAIHSTHSVTLYDCHSMKGNOLMKYRKDKTLYPHPVGSGSCMDSESHPRIEMNT 4800
Oy	578	CNPSSLTQQMLFEHTNSTVLKEFNEN 603
Dd	481	CNPSSLTQQMLFEHTNSTVLKEFNEN 506
<b>RESULT 6</b>		
AAAG79781	standard; protein; 506 AA.	
XX	AAAG79781;	
AC	AAAG79781;	
XX	16-APR-2003 (first entry)	
DT	Carbohydrate-associated protein (CHOP) -3.	
DE		
XX	Human; carbohydrate associated polypeptides; CHOP; cancer;	
KW	carbohydrate metabolism; diabetes; anaemia; hypoglycaemia; obesity;	
KW	glycogen storage disease; neurological dysfunction; gene therapy;	
KW	cell proliferation; actinic keratosis; arteriosclerosis; inflammation;	
KW	atherosclerosis; cirrhosis; hepatitis; psoriasis; autoimmune disorder;	
KW	acquired immunodeficiency syndrome; AIDS; humanized; transgenic;	
KW	Addison's disease; allergy; asthma; contact dermatitis; bronchitis; gout;	
KW	Hashimoto's thyroiditis; infection; reproduction; endometriosis;	
KW	polycystic ovary syndrome; ovarian hyperstimulation syndrome;	
KW	Down syndrome; cystic fibrosis; sickle cell anemia; thalassemia;	
KW	myocarditis; cardiomyopathy; Alzheimer's disease; Parkinson's disease;	
KW	dementia; depression; epilepsy; Tourette's disorder; schizophrenia;	
KW	central nervous system; cerebral palsy; mood; anxiety; knockin.	
XX		
OS	Homo sapiens.	
XX		
FH	Key	Location/Qualifiers
FT	Peptide	3..503
FT	/label= Acetylgalactosaminyl transferase	
FT	/note= "Identified by BLAST_DOMO"	
FT	Peptide	8..493
FT	/label= Acetylgalactosaminyl transferase	
FT	/note= "Identified by BLAST_DOMO"	
FT	Peptide	11..492
FT	/label= Acetylgalactosaminyl transferase	
FT	/note= "Identified by BLAST_DOMO"	
FT	Peptide	51..236
FT	/label= Glycosyl transferase sequence	
FT	/note= "Identified by HMMER_PFBM"	
FT	Peptide	202..362
FT	/label= N-acetylglactosaminyl transferase	
FT	/note= "Identified by BLAST_PRODOW"	
FT	Region	364..402
FT	/label= QXW lectin repeat	
FT	/note= "Identified by HMMER_PFBM"	
FT	Region	417..455
FT	/label= QXW lectin repeat	
FT	/note= "Identified by HMMER_PFBM"	
FT	Region	456..495
FT	/label= QXW lectin repeat	
FT	/note= "Identified by HMMER_PFBM"	
XX		

PN WO200297060-A2.  
 XX 05-DEC-2002.  
 XX 22-MAY-2002; 2002WO-US018354.  
 XX 25-MAY-2001; 2001US-0293768P.  
 XX 01-AUG-2001; 2001US-0309548P.  
 XX 23-AUG-2001; 2001US-0314400P.  
 XX 19-OCT-2001; 2001US-0343706P.  
 XX 07-DEC-2001; 2001US-0337999P.  
 XX (INCYTE GENOMICS INC.  
 XX Swarnakar A, Gorvad AE, Hafalia AJA, Duggan BM, Emerling BM,  
 XX Ison CH, Nguyen DB, Lee EA, Yue H, Foreyche IJ, Li JX,  
 XX Thangavelu K, Walla NK, Burford N, Mason PM, Lal PG, Lee S,  
 XX Becha SD, Tang YT;  
 XX WPI: 2003-140462/13.  
 XX N-PSDB; ABA00833.  
 PT Novel human carbohydrate associated polypeptide, useful in diagnosis,  
 PT treatment and prevention of carbohydrate metabolism, cell proliferative,  
 PT autoimmune/inflammatory, reproductive, and neurological disorders.  
 PS Claim 1; Page 126-27; 141pp; English.  
 XX The sequences given in AAG79779-88 represent human carbohydrate  
 XX associated polypeptides (CHOP). The CHOP polypeptides and the nucleotide  
 XX sequences encoding them, are useful for diagnosing, treating and  
 XX preventing cancer, carbohydrate metabolic disorders (e.g. diabetes,  
 XX anaemia, hypoglycaemia, obesity, glycogen storage disease, neurological  
 XX dysfunctions), cell proliferative disorders (e.g. actinic keratosis,  
 XX arteriosclerosis, atherosclerosis, cirrhosis, hepatitis, psoriasis),  
 XX autoimmune/inflammatory disorders (such as acquired immunodeficiency  
 XX syndrome (AIDS), Addison's disease, allergies, asthma, contact  
 XX dermatitis, bronchitis, gout, Hashimoto's thyroiditis), viral, bacterial,  
 XX fungal, parasitic, protozoal and helminthic infections, reproductive  
 XX disorders (e.g. endometriosis, polycystic ovary syndrome, ovarian  
 XX hyperstimulation syndrome), genetic disorders (e.g. Down syndrome, cystic  
 XX fibrosis, sickle cell anaemia, thalassemia), cardiac disorders (e.g.  
 XX myocarditis, cardiomyopathy), neurological disorders (e.g. Alzheimer's  
 XX disease, Parkinson's disease, dementia, depression, epilepsy, Tourette's  
 XX disorder, schizophrenia), and developmental disorders of central nervous  
 XX system (e.g. cerebral palsy), mental disorders (e.g. mood, anxiety). CHOP  
 XX proteins are useful in a number of drug screening techniques, and to  
 XX analyse the proteome of a tissue or cell type. CHOP cDNA is useful for  
 XX creating "knockin" humanized animals or transgenic animals to model human  
 XX diseases, in somatic or germ-line gene therapy, to generate a transcript  
 XX image of a tissue or cell type, for detecting differences in the  
 XX chromosomal location due to translocation, inversion, etc., among normal,  
 XX carrier or affected individuals, and as hybridization probes for mapping  
 XX naturally occurring genomic sequences  
 SQ Sequence 506 AA;  
 Query Match 84.5%; Score 2771; DB 6; Length 506;  
 Best Local Similarity 100.0%; Pred. No. 8.2e-274; Indels 0; Gaps 0;  
 Matches 506; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 98 MTDARVQAVRENGFNIVYSDKISLNSLPDIRHPCNSKRYLETLPNTSIIIPFHNHG 157  
 DB 1 MTDARVQAVRENGFNIVYSDKISLNSLPDIRHPCNSKRYLETLPNTSIIIPFHNHG 60  
 QY 158 WSLIRTVYASVYLNNSPPVLAIVLVDDPSDREHKKLEEDYMALEPSVRLIRTKRGEL 217  
 DB 61 WSLIRTVYASVYLNNSPPVLAIVLVDDPSDREHKKLEEDYMALEPSVRLIRTKRGEL 120  
 QY 218 IRTMVLGASVATGVDTFLDHSCEANVMWLPILDIRIRNRTYICPMIDIVYDHDPEYE 277  
 DB 121 IRTMVLGASVATGVDTFLDHSCEANVMWLPILDIRIRNRTYICPMIDIVYDHDPEYE 180

QY 278 TQAGDANRGAIPDMEMYYKRIPIPELOKADPSDPFSPVWAGGIFAVDRKMFHMLGGYDP 337  
 DB 181 TQAGDANRGAIPDMEMYYKRIPIPELOKADPSDPFSPVWAGGIFAVDRKMFHMLGGYDP 240  
 QY 338 GLRTWGEQYEISKVMWCGGRMBDIPCSRVGHYRYKYVYKVPAGVSLRNLKRVAEVW 397  
 DB 241 GLRTWGEQYEISKVMWCGGRMBDIPCSRVGHYRYKYVYKVPAGVSLRNLKRVAEVW 300  
 QY 398 MDEYAEYIYQRRPEYRLSAGDVAVQKKLRSLNCKSFKKFMTKLIANDLPKFPVPEPA 457  
 DB 301 MDEYAEYIYQRRPEYRLSAGDVAVQKKLRSLNCKSFKKFMTKLIANDLPKFPVPEPA 360  
 QY 458 AAMGEIRNVGTGLCADTRKQALGSPRLBECVGRGSAANNNOVFPTREDIRPDPO 517  
 DB 361 AAMGEIRNVGTGLCADTRKQALGSPRLBECVGRGSAANNNOVFPTREDIRPDPO 420  
 QY 518 HTKKECFDAISHSPVLYDCHSKMGQMKYRKDKTLVYPSGSCDSCSDHRIEMNT 577  
 DB 421 HTKKECFDAISHSPVLYDCHSKMGQMKYRKDKTLVYPSGSCDSCSDHRIEMNT 480  
 QY 578 CNPSSLTQOMLFHEHTNSTVLEKFNRN 603  
 DB 481 CNPSSLTQOMLFHEHTNSTVLEKFNRN 506  
 RESULT 7  
 ID AAU07771 standard; protein; 535 AA.  
 AC AAU07771;  
 XX 04-DEC-2001 (first entry)  
 DT 04-DEC-2001 (first entry)  
 XX Human novel transferase protein, NHP #14.  
 DE Human novel transferase protein, NHP #14.  
 XX Human, transferase, breast cancer; prostate cancer; immunogen;  
 KM gene therapy; antisense.  
 XX Homo sapiens.  
 OS Homo sapiens.  
 PN WO200164903-A2.  
 XX 07-SEP-2001.  
 PD 28-FEB-2001; 2001WO-US006460.  
 PF 29-FEB-2000; 2000US-0185920P.  
 PR 02-MAR-2000; 2000US-0186558P.  
 PR 24-MAR-2000; 2000US-0191849P.  
 XX (LEXI-) LEXICON GENETICS INC.  
 XX Donoho G, Hillman E, Turner CA, Friedrich G, Abuln A;  
 PI Zambrowicz B, Sande AT, Walke DW, Wilganowski NL, Hu Y, Kleke JA,  
 PI Potter DG;  
 XX WPI: 2001-550185/61.  
 DR N-PSDB; AAS12618.  
 PT Novel nucleic acid sequences encoding novel human proteins useful for  
 PT diagnosis, drug screening, clinical trial monitoring and treatment of  
 PT diseases and disorders.  
 PS Claim 7; Page 45-46; 60pp; English.  
 XX The invention relates to isolated nucleic acids encoding novel human  
 XX transferase proteins (NHP). The nucleic acids and proteins are useful for  
 XX diagnosis, drug screening, clinical trial monitoring and treatment of  
 XX diseases and disorders e.g. breast and prostate cancer. NHPs can also be  
 XX useful for augmenting the efficacy of chemotherapeutic agents used in  
 XX treatment of breast or prostate cancer. The nucleic acid is also useful  
 XX in NHP gene regulation, and as antisense primers in amplification  
 XX reactions of NHP gene sequences. NHPs are useful for producing

CC antibodies. The present sequence represents a novel human transferase  
CC which has sequence similarity to N-acetyl-galactosaminyltransferase  
XX  
SQ Sequence 535 AA;

Query Match 83.8%; Score 2746.5; DB 4; Length 535;  
Best Local Similarity 94.6%; Pred. No. 2,96-271;  
Matches 506; Conservative 0; Mismatches 0; Indels 29; Gaps 1;

QY 98 MTDNERDQAYRENGFNIVYSDKISLRSLPDIDHPNCSKRYLETLPNTSIIIPFNEG 157  
DB 1 MTDNERDQAYRENGFNIVYSDKISLRSLPDIDHPNCSKRYLETLPNTSIIIPFNEG 60  
QY 158 WSSLLRTHSVLNSPPELVAEIVLVDFSDREHLKKPLBRYMLPFSVRLTRKKEGL 217  
DB 61 WSSLLRTHSVLNSPPELVAEIVLVDFSDREHLKKPLBRYMLPFSVRLTRKKEGL 120  
QY 218 IRTTMLGASVATGVDITFLDSHCANVMPLPLDRIARNKTIYCEMIDVIDHDDEPRYE 277  
DB 121 IRTTMLGASVATGVDITFLDSHCANVMPLPLDRIARNKTIYCEMIDVIDHDDEPRYE 180  
QY 278 TQAGDANRGAAPDWMYTKRIPIPELOKADSPDPESPVAAGLFAVDKMFELGGYDP 337  
DB 181 TQAGDANRGAAPDWMYTKRIPIPELOKADSPDPESPVAAGLFAVDKMFELGGYDP 240  
QY 338 GLEIWGGEQYEISFK-----YVMCGRMEDIPCSRV 368  
DB 241 GLEIWGGEQYEISFKGLHMLPRVLSNSMPQAVFLPRAPNMLALQVMMCGRMEDIPCSRV 300  
QY 369 GHYRKVPKVPKPAVSLAENLKRAVEMWDEVAEYIYQRPPEYRHLISAGDVAVQKRLS 428  
DB 301 GHYRKVPKVPKPAVSLAENLKRAVEMWDEVAEYIYQRPPEYRHLISAGDVAVQKRLS 360  
QY 429 SLNCKSFKFMFTKIAMDLPKFYPPVEBPAAAGSIRNVGTGLCADTKHGLGSPRLGEC 488  
DB 361 SLNCKSFKFMFTKIAMDLPKFYPPVEBPAAAGSIRNVGTGLCADTKHGLGSPRLGEC 420  
QY 489 VRRGEAAMNNMVOFTTWREDIRPQDPOHTKKCFPAISHTSPVTLYDCHSMKGNLWK 548  
DB 421 VRRGEAAMNNMVOFTTWREDIRPQDPOHTKKCFPAISHTSPVTLYDCHSMKGNLWK 480  
QY 549 YRKDKTLVHPVSGSCMDCSDSDHRIFMNTCNPSLSLTQOVLFEHTNSTVLEKFNEN 603  
DB 481 YRKDKTLVHPVSGSCMDCSDSDHRIFMNTCNPSLSLTQOVLFEHTNSTVLEKFNEN 535

RESULT 8  
ID AAB94733 standard; protein; 339 AA.  
XX  
AC AAB94733;  
XX  
DT 26-JUN-2001 (first entry)  
XX  
DE Human protein sequence SEQ ID NO:15766.  
XX  
KM Human; primer; detection; diagnosis; antinease therapy; gene therapy.  
XX  
OS Homo sapiens.  
XX  
PN BP1074617-AA2.  
XX  
PD 07-FEB-2001.  
XX  
PF 28-JUL-2000; 2000BP-00116126.  
XX  
PR 29-JUL-1999; 99JB-00248036.  
XX  
PR 27-AUG-1999; 99JB-00300253.  
XX  
PR 11-JAN-2000; 2000JP-00118776.  
XX  
PR 02-MAY-2000; 2000JP-00183767.  
XX  
PR 09-JUN-2000; 2000JP-00241899.  
XX  
PA (HELI-) HELIX RES INST.

XX Oca T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
XX WPI: 2001-318749/34.  
XX

PT Primer sets for synthesizing polynucleotides, particularly the 5602 full-  
PT length cDNAs defined in the specification, and for the detection and/or  
PT diagnosis of the abnormality of the proteins encoded by the full-length  
PT cDNAs.

PS Claim 8; SEQ ID NO 15766; 2537pp + Sequence Listing; English.

CC The present invention describes primer sets for synthesizing 5602 full-  
CC length cDNAs defined in the specification. Where a primer set comprises:  
CC (a) an oligo-dT primer and an oligonucleotide complementary to the  
CC complementary strand of a polynucleotide which comprises one of the 5602  
CC nucleotide sequences defined in the specification, where the  
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
CC of an oligonucleotide comprising a sequence complementary to the  
CC complementary strand of a polynucleotide which comprises a 5'-end  
CC sequence and an oligonucleotide comprising a sequence complementary to a  
CC polynucleotide which comprises a 3'-end sequence, where the  
CC oligonucleotide comprises at least 15 nucleotides and the combination of  
CC the 5'-end sequence/3'-end sequence is selected from those defined in the  
CC specification. The primer sets can be used in antisense therapy and in  
CC gene therapy. The primers are useful for synthesizing polynucleotides,  
CC particularly full-length cDNAs. The primers are also useful for the  
CC detection and/or diagnosis of the abnormality of the proteins encoded by  
CC the full-length cDNAs. The primers allow obtaining of the full-length  
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893  
CC represent human amino acid sequences; and AAH13629 to AAH13632 represent  
CC oligonucleotides, all of which are used in the exemplification of the  
CC present invention

XX SQ Sequence 339 AA;

Query Match 57.7%; Score 1893; DB 4; Length 339;  
Best Local Similarity 99.7%; Pred. No. 3,1e-184;  
Matches 338; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 265 MIDVIDHDDPRYETQGDANRGAAPDWMYTKRIPIPELOKADSPDPESPVAAGLFAV 324  
DB 1 MIDVIDHDDPRYETQGDANRGAAPDWMYTKRIPIPELOKADSPDPESPVAAGLFAV 60  
QY 325 DRKMFELGGYDPGLIHWGGEQYEISFKVMMCGRMEDIPCSRVGHIYRKVPYKPAV 384  
DB 61 DRKMFELGGYDPGLIHWGGEQYEISFKVMMCGRMEDIPCSRVGHIYRKVPYKPAV 120  
QY 385 SLARNLKRAVEMWDEVAEYIYQRPPEYRHLISAGDVAVQKRLSSLNCKSFKFMFTKIAM 444  
DB 121 SLARNLKRAVEMWDEVAEYIYQRPPEYRHLISAGDVAVQKRLSSLNCKSFKFMFTKIAM 180  
QY 445 DLKFPYPPVBPAAAGSIRNVGTGLCADTKHGLGSPRLDECVRRGSAAMNNMVOFT 504  
DB 181 DLKFPYPPVBPAAAGSIRNVGTGLCADTKHGLGSPRLDECVRRGSAAMNNMVOFT 240  
QY 505 FTWRREDIRPQDPOHTKKCFDAISHTSPVTLYDCHSMKGNLWKRYRKDKTLVHPVSGSCM 564  
DB 241 FTWRREDIRPQDPOHTKKCFDAISHTSPVTLYDCHSMKGNLWKRYRKDKTLVHPVSGSCM 300  
QY 565 DCSDSDHRIFMNTCNPSLSLTQOVLFEHTNSTVLEKFNEN 603  
DB 301 DCSDSDHRIFMNTCNPSLSLTQOVLFEHTNSTVLEKFNEN 339

RESULT 9  
ID AAU07776 standard; protein; 366 AA.  
XX  
AC AAU07776;  
XX

DT 04-DEC-2001 (first entry)  
 XX Human novel transferase protein, NHP #19.  
 DE Human; transferase; breast cancer; prostate cancer; immunogen;  
 XX gene therapy; anticense.  
 KM Homo sapiens.  
 XX WO200164903-A2.  
 PN 07-SEP-2001.  
 PD 28-FEB-2001; 2001WO-US006460.  
 XX 29-FEB-2000; 2000US-0185920P.  
 PF 02-MAR-2000; 2000US-018558P.  
 PR 24-MAR-2000; 2000US-0191849P.  
 PA (LEXI-) LEXICON GENETICS INC.  
 XX Donoho G, Hillman E, Turner CA, Friedrich G, Abuln A;  
 PI Zambrowicz B, Sands AT, Walke DW, Wilganowski NL, Hu Y, Kieke JA;  
 PI Potter DG;  
 XX WPI; 2001-550185/61.  
 DR N-PSDB; AAS12623.  
 XX Novel nucleic acid sequences encoding novel human proteins useful for  
 PT diagnosis, drug screening, clinical trial monitoring and treatment of  
 PT diseases and disorders.  
 XX Disclosure; Page 52-53; 60pp; English.  
 PS The invention relates to isolated nucleic acids encoding novel human  
 CC transferase proteins (NHP). The nucleic acids and proteins are useful for  
 CC diagnosis, drug screening, clinical trial monitoring and treatment of  
 CC diseases and disorders e.g. breast and prostate cancer. NHPs can also be  
 CC useful for augmenting the efficacy of chemotherapeutic agents used in  
 CC treatment of breast or prostate cancer. The nucleic acid is also useful  
 CC in NHP gene regulation, and as antisense primers in amplification  
 CC reactions of NHP gene sequences. NHPs are useful for producing  
 CC antibodies. The present sequence represents a novel human transferase  
 CC which has sequence similarity to N-acetyl-Galactosaminyltransferase  
 CC  
 XX Sequence 366 AA;  
 SQ  
 Query Match 57.3%; Score 1877; DB 4; Length 366;  
 Best Local Similarity 100.0%; Pred. No. 1.5e-182;  
 Matches 353; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MRRREKRLQVALVALVALVLPVNGMALYREROPDGTGGSGAAVAPAGQSGSHSRQK 60  
 DB 1 MRRREKRLQVALVALVALVLPVNGMALYREROPDGTGGSGAAVAPAGQSGSHSRQK 60  
 QY 61 KTFPLGDOGQKLMDHDKAIRDADQVNGEGREPYMTDAERVDQAVRENGFNIVYSDK 120  
 DB 61 KTFPLGDOGQKLMDHDKAIRDADQVNGEGREPYMTDAERVDQAVRENGFNIVYSDK 120  
 QY 121 ISLNRSLPDIRHPNCSKRYETLPNTSIIIPFNHGGSSLLRTVHSLVNSPPPLVAEI 180  
 DB 121 ISLNRSLPDIRHPNCSKRYETLPNTSIIIPFNHGGSSLLRTVHSLVNSPPPLVAEI 180  
 QY 121 ISLNRSLPDIRHPNCSKRYETLPNTSIIIPFNHGGSSLLRTVHSLVNSPPPLVAEI 180  
 DB 121 ISLNRSLPDIRHPNCSKRYETLPNTSIIIPFNHGGSSLLRTVHSLVNSPPPLVAEI 180  
 QY 181 VLVDPSREHLKPLEDYMALFESVRLTKRKGRLRTKMLGASVATGVTFLDSHC 240  
 DB 181 VLVDPSREHLKPLEDYMALFESVRLTKRKGRLRTKMLGASVATGVTFLDSHC 240  
 QY 181 VLVDPSREHLKPLEDYMALFESVRLTKRKGRLRTKMLGASVATGVTFLDSHC 240  
 DB 181 VLVDPSREHLKPLEDYMALFESVRLTKRKGRLRTKMLGASVATGVTFLDSHC 240  
 QY 241 EAVNWLPEPLLDRIARNKTIIVCPMIDVIDHDFRYETQAGDAMKAGFDMWYTKRIP 300  
 DB 241 EAVNWLPEPLLDRIARNKTIIVCPMIDVIDHDFRYETQAGDAMKAGFDMWYTKRIP 300  
 QY 241 EAVNWLPEPLLDRIARNKTIIVCPMIDVIDHDFRYETQAGDAMKAGFDMWYTKRIP 300  
 DB 241 EAVNWLPEPLLDRIARNKTIIVCPMIDVIDHDFRYETQAGDAMKAGFDMWYTKRIP 300  
 QY 301 PELQKADSDPFPESPVMAGLFAVDRKMFELGSDPGLTGWGEQYEISERV 353  
 DB 301 PELQKADSDPFPESPVMAGLFAVDRKMFELGSDPGLTGWGEQYEISERV 353

DB 301 PELQKADSDPFPESPVMAGLFAVDRKMFELGSDPGLTGWGEQYEISERV 353  
 RESULT 10  
 ID AAG62600  
 XX AAG62600 standard; protein; 407 AA.  
 AC AAG62600;  
 XX 06-SEP-2001 (first entry)  
 DT Human acetyl galactosyl transferase 45.  
 XX Human acetyl galactosyl transferase 45; GalNAc-T45; cancer; haemopathy;  
 DE Human; acetyl galactosyl transferase 45; GalNAc-T45; cancer; haemopathy;  
 KM HIV infection; immunological disease; inflammation; gene therapy.  
 XX Homo sapiens.  
 XX WO200138545-A1.  
 PN 31-MAY-2001.  
 PD 20-NOV-2000; 2000MO-CN000473.  
 XX 24-NOV-1999; 99CN-00124100.  
 PR (BIOR-) BIOROAD GENE DEV LTD SHANGHAI.  
 PA Mao Y, Xie Y;  
 PI WPI; 2001-355942/37.  
 DR N-PSDB; AAH45701.  
 XX Human acetyl galactosyl transferase 45 and encoded polynucleotide, used  
 PT in diagnosis and treatment of malignant tumors, hemopathy, human  
 PT immunodeficiency virus infection, immunological diseases and  
 PT inflammation.  
 XX Claim 1; Page 20-21; 33pp; Chinese.  
 PS The present invention provides the protein and coding sequences of the  
 CC human acetyl galactosyl transferase 45 (GalNAc-T45). The sequences can be  
 CC used in the treatment of cancer, haemopathy, HIV infection, immunological  
 CC diseases and inflammation. The present sequence is the protein of the  
 CC invention  
 CC  
 XX Sequence 407 AA;  
 SQ  
 Query Match 55.5%; Score 1818; DB 4; Length 407;  
 Best Local Similarity 78.5%; Pred. No. 2e-176;  
 Matches 317; Conservative 36; Mismatches 51; Indels 0; Gaps 0;  
 QY 200 MALPFSVRLTKRKGRLRTKMLGASVATGVTFLDSHC EAVNWLPEPLLDRIARNK 259  
 DB 1 MARFSKRLVATKRRKGRLRTKMLGASVATGVTFLDSHC EAVNWLPEPLLDRIARNK 60  
 QY 260 TIVCPMIDVIDHDFRYETQAGDAMKAGFDMWYTKRIPPELQKADSDPFPESPVMAG 319  
 DB 61 TIVCPMIDVIDHDFRYETQAGDAMKAGFDMWYTKRIPPELQKADSDPFPESPVMAG 120  
 QY 320 GLFAVDRKMFELGSDPGLTGWGEQYEISFKVMCGGEMFDPVCSRVGHIYRKIVPYK 379  
 DB 121 GLFAVDRKMFELGSDPGLTGWGEQYEISFKVMCGGEMFDPVCSRVGHIYRKIVPYK 180  
 QY 380 VPAGVSLARNLKRAEYVWMBEYAEIYORPREYHLSAGDVAQKXLRSSLNCKSPKWM 439  
 DB 181 VPAGVSLARNLKRAEYVWMBEYAEIYORPREYHLSAGDVAQKXLRSSLNCKSPKWM 240  
 QY 440 TKIAMDLPKFPYPPVPAAGGEIRNYGTGLCADTKHAGLSPLRLEGCVRGGEAAMNN 499  
 DB 241 AAVAMDPKYYPPVPPPAAGGEIRNYAANLCVDSKKGAGCTELRLDIQVDSGSEKTMH 300  
 QY 500 MOVFTFWREDIRGDPQHTKKFCFDALSHTSPTLYDCHSMKGNQLMKYRKDKTLVHPV 559

Db 301 EQFTFMRDIDRGEELHTRKFCFDALSHNSPVTLYDCHGMKNQJMGARKORTLHPV 360  
 QY 560 SSGCMDCSESDHRIFMNTCNPSLITQOMLFPHNSTYLEKFNKN 603  
 Db 361 SNSCMDCNPAAKKIFMARCDPLSETQOMIEFHIMTVLEKFNH 404

## RESULT 11

ABG08187  
 ID ABG08187 standard; protein; 319 AA.  
 XX  
 AC ABG08187;  
 XX  
 DT 13-FEB-2002 (first entry)  
 XX  
 DE Novel human diagnostic protein #8178.  
 XX  
 KM Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 KM food supplement; medical imaging; diagnostic; genetic disorder.  
 XX  
 OS Homo sapiens.  
 XX  
 PN MO200175067-A2.  
 XX  
 PD 11-OCT-2001.  
 XX  
 PF 30-MAR-2001; 2001MO-US008631.  
 XX  
 PR 31-MAR-2000; 2000US-00540217.  
 PR 23-AUG-2000; 2000US-00649167.  
 XX  
 PA (HYSE-) HYSEQ INC.  
 XX  
 PI Drmanac RT, Liu C, Tang YT;  
 XX  
 DR WPI: 2001-639362/73.  
 DR N-PSDB; AAS72374.  
 XX  
 PT New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity.

Claim 20; SEQ ID NO 38546; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and polypeptide (II)  
 CC sequences. (I) is useful as hybridisation probes, polymerase chain  
 CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,  
 CC and in recombinant production of (II). The polynucleotides are also used  
 CC in diagnostics as expressed sequence tags for identifying expressed  
 CC genes. (I) is useful in gene therapy techniques to restore normal  
 CC activity of (II) or to treat disease states involving (II). (II) is  
 CC useful for generating antibodies against it, detecting or quantitating a  
 CC polypeptide in tissue, as molecular weight markers and as a food  
 CC supplement. (II) and its binding partners are useful in medical imaging  
 CC of sites expressing (II). (I) and (II) are useful for treating disorders  
 CC involving aberrant protein expression or biological activity. The  
 CC polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic  
 CC amino acid sequences of the invention. Note: The sequence data for this  
 CC patent did not appear in the printed specification, but was obtained in  
 CC electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences

XX Sequence 319 AA;

Query Match 42.8%; Score 1403.5; DB 4; Length 319;  
 Best Local Similarity 90.8%; Pred. No. 3.5e-134;  
 Matches 258; Conservative 4; Mismatches 19; Indels 3; Gaps 2;

QY 322 FAVDRKFWELGEGYDPEGLBIWGEQYELSPKVMCGGRMEDIPCSRIGHIYRKVYPKVP 381  
 Db 12 FARCAKQWPE-GGMDLEIAETSSGLFLAMQVMCGGRMEDIPCSRIGHIYRKVYPKVP 70  
 QY 382 AGVSLAR--NLAKVAEVMDEVAEYIYQRRPEYRHLSAGDVAVQKLRSSLNCKSPKFM 439  
 Db 71 AGVSLARSKLKKVAVVMDEVAEYIYQRRPEYRHLSAGDVAVQKLRSSLNCKSPKFM 130  
 QY 440 TKIAMDLPKFPYPPVEPPAAWGEIRNVGTGLCADTKHAGLSPRLREGYRGEAAWNN 499  
 Db 131 TKIAMDLPKFPYPPVEPPAAWGEIRNVGTGLCADTKHAGLSPRLREGYRGEAAWNN 190  
 QY 500 MOVFTFTMRDIDRGEELHTRKFCFDALSHNSPVTLYDCHGMKNQJMGARKORTLHPV 559  
 Db 191 MOVFTFTMRDIDRGEELHTRKFCFDALSHNSPVTLYDCHGMKNQJMGARKORTLHPV 250  
 QY 560 SSGCMDCSESDHRIFMNTCNPSLITQOMLFPHNSTYLEKFNKN 603  
 Db 251 SSGCMDCSESDHRIFMNTCNPSLITQOMLFPHNSTYLEKFNKN 294

## RESULT 12

ADE25865  
 ID ADE25865 standard; protein; 275 AA.  
 XX  
 AC ADE25865;  
 XX  
 DT 29-JAN-2004 (first entry)  
 XX  
 DE Human N-acetylglucosaminyl transferase T10.

XX cytostatic; respiratory; antiasthmatic; antiinflammatory; endocrine; CNS;  
 KM immunosuppressive; antiarthritic; antibacterial; vulnery;  
 KM GalNAc-transferase lectin domain inhibitor;  
 KM N-acetylglucosaminyl transferase; GalNAc-transferase; tumour; cancer;  
 KM GalNAc-T7; GalNAc-T2; GalNAc-T3; GalNAc-transferase; GalNAc-T4;  
 KM lung disease; mucous accumulation; asthma; chronic bronchitis;  
 KM smoker's lung; cystic fibrosis; exocrine gland disease;  
 KM mucin secretion decrease; Sjogren's syndrome; dry mouth;  
 KM selection-mediated leukocyte trafficking; autoimmunity; arthritis;  
 KM leukaemia; lymphoma; immunosuppression; sepsis; wound healing;  
 KM inflammation; mucin secretion inhibition; hypersecretion;  
 KM mucin accumulation; chronic obstructive respiratory pulmonary disease;  
 KM asthma; lectin domain inhibition; GalNAc-glycopeptide;  
 KM Galbeta1-3GalNAc-glycopeptide; UDP-Gal donor; glycosyltransferase;  
 KM core 1 beta1,3-galactosyltransferase; alpha2,6-sialyltransferase;  
 KM N-acetylglucosaminyl transferase 4Ap24 glycopeptide; O-glycosylation;  
 KM breast cancer; human; GalNAcT10.

XX Homo sapiens.  
 OS  
 PN US2003186850-A1.  
 XX  
 PD 02-OCT-2003.  
 XX  
 PF 12-NOV-2002; 2002US-00292896.  
 XX  
 PR 11-MAY-2000; 2000US-0203311P.  
 PR 10-MAY-2001; 2001MO-DR000328.  
 PR 08-NOV-2002; 2002US-0425204P.  
 XX  
 PA (GLYC-) GLYC02YW APS.  
 XX  
 PI Clausen H, Bennett EP, Hasean H, Reis CA;  
 XX  
 DR WPI: 2003-831259/77.  
 XX  
 PT Modulating N-acetylglucosaminyltransferase functions comprises  
 PT administration of an N-acetylglucosaminyltransferase lectin domain  
 PT inhibitor.

PS Disclosure; Page 11; 65pp; English.

XX The invention describes a method of modulating polypeptide N-  
CC acetylglucosaminyl (GALNAc)-transferase functions comprising  
CC administration of a GALNAc-transferase lectin domain inhibitor (I). (1)  
CC is used for modulating functions of GALNAc-transferase (e.g. GALNAc-T4,  
CC GALNAc-T7, GALNAc-T2, and GALNAc-T3), for inhibiting at least one lectin  
CC domain of the GALNAc-transferase and modulating the function mediated by  
CC the domain for preparing a medicament for treating tumours and cancers,  
CC lung diseases associated with mucous accumulation (e.g. asthma, chronic  
CC bronchitis, smoker's lung, and cystic fibrosis), diseases of exocrine  
CC glands associated with increased or decreased mucin secretion (e.g.  
CC Sjogren's syndrome and dry mouth), disorders associated with  
CC dysregulation of selectin-mediated leukocyte trafficking (e.g.  
CC autoimmunity, arthritis, leukemias, lymphomas, immunosuppression,  
CC sepsis, wound healing, acute and chronic inflammation); inhibiting mucin  
CC secretion; and inhibiting hypersecretion and accumulation of mucin in the  
CC lungs of mammals (e.g. humans) suffering from chronic obstructive  
CC respiratory pulmonary diseases, asthma, and cystic fibrosis. (1) is  
CC effective in inhibiting at least one lectin domain of the GALNAc-  
CC transferase and modulating the inhibited function mediated by the domain  
CC (preferably GALNAc-glycopeptide, galbetail-3GALNAc-glycopeptide, or  
CC peptide specificity of the transferase, or UDP-gal donor substrate  
CC specificity). (1) selectively inhibits at least one member of the GALNAc-  
CC transferase family without inhibiting other glycosyltransferases  
CC (preferably core 1 beta1,3-galactosyltransferase, alpha2,6-  
CC sialyltransferase, or glycosyltransferase functioning in the O-  
CC glycosylation pathway). This is the amino acid sequence of human N-  
CC acetylglucosaminyl transferase T10.  
XX

Sequence 275 AA;

Query Match 42.6%; Score 1396; DB 7; Length 275;

Best Local Similarity 93.7%; Pred. No. 1.6e-133;

Matches 253; Conservative 4; Mismatches 11; Indels 2; Gaps 1;

QY 336 DPLGELMGGEYYSFKYWMCGRMEDIPCSRVGHIYRKYVPYKYPAGVSLA--RNLRKV 393  
DB 6 DGELEETSSSLFLAMQVMCGRMEDIPCSRVGHIYRKYVPYKYPAGVSLARVYTLKRV 65  
QY 394 AEVMMDEVAEYIYORRPREYRLHLSAGDVAVOKKRLSSLNCKSPKWMPTIAMDLPKYPVPV 453  
DB 66 AEVMMDEVAEYIYORRPREYRLHLSAGDVAVOKKRLSSLNCKSPKWMPTIAMDLPKYPVPV 125  
QY 454 EPPAAAMEIRNVGTGLCADTKHGALGSPRLRLGECVGRGEAAMNNQVFTFTWREDIRP 513  
DB 126 EPPAAAMEIRNVGTGLCADTKHGALGSPRLRLGECVGRGEAAMNNQVFTFTWREDIRP 185  
QY 514 GDPQHTKKFCPDALSHTSPTVLYDCHSMKGNQLMKTKRKDKTLVHPVSGSCMDCSSESHRI 573  
DB 186 GDPQHTKKFCPDALSHTSPTVLYDCHSMKGNQLMKTKRKDKTLVHPVSGSCMDCSSESHRI 245  
QY 574 FMNTCNPSLTQOWLFETHNSTVLEKFNRN 603  
DB 246 FMNTCNPSLTQOWLFETHNSTVLEKFNRN 275

RESULT 13

ADQ17613 standard; protein, 276 AA.

ADQ17613;

26-AUG-2004 (first entry)

Human soft tissue sarcoma-upregulated protein - SEQ ID 430.

soft tissue sarcoma; cytosolatic; gene therapy; vaccine; screening; human.

Homo sapiens.

WC02004048938-A2.

10-JUN-2004.

XX 26-NOV-2003; 2003WO-US038193.  
XX 26-NOV-2002; 2002US-0429739P.  
XX (PROT-) PROTEIN DESIGN LABS INC.  
XX Aziz N, Gineburg WM, Zlotnik A;  
XX WPI; 2004-441206/41.  
XX

PT Early detection of soft tissue sarcoma comprises determining expression  
PT of a gene in a first soft tissue sample and a normal soft tissue sample  
PT and comparing the gene expression, also useful in treating soft tissue  
PT sarcoma.

Example 2; SEQ ID NO 430; 210pp; English.

XX The invention relates to a novel method for detecting soft tissue sarcoma  
XX which comprises obtaining a first soft tissue sample from an individual  
XX and a normal soft tissue sample from the same or different individual,  
XX determining the expression of a gene in both samples and comparing the  
XX expression of the gene in both soft tissue samples, where a higher level  
XX of protein expression in the first soft tissue sample indicates the  
XX presence of soft tissue sarcoma. The method of the invention has  
XX cytosolatic applications and may be useful for detecting soft tissue  
XX sarcoma, possibly via gene therapy or vaccine production. The nucleic  
XX acid sequences may be useful in diagnostic and screening applications.  
XX The current sequence is that of a human soft tissue sarcoma-upregulated  
XX protein of the invention. The current sequence is not shown within the  
XX specification per se but was submitted in CD format by the inventor.

Sequence 276 AA;

Query Match 42.5%; Score 1392; DB 8; Length 276;

Best Local Similarity 97.3%; Pred. No. 4.2e-133;

Matches 250; Conservative 3; Mismatches 2; Indels 2; Gaps 1;

QY 349 ISFKVMCGRMEDIPCSRVGHIYRKYVPYKYPAGVSLA--RNLRKVAEVMDEVAEYIY 406  
DB 20 LAMQVMCGRMEDIPCSRVGHIYRKYVPYKYPAGVSLARVYTLKRVAEVMMDEVAEYIY 79  
QY 407 QRRPREYRLHLSAGDVAVOKKRLSSLNCKSPKWMPTIAMDLPKYPVPVPEPPAAAMEIRNV 466  
DB 80 QRRPREYRLHLSAGDVAVOKKRLSSLNCKSPKWMPTIAMDLPKYPVPVPEPPAAAMEIRNV 139  
QY 467 GTGLCADTKHGALGSPRLRLGECVGRGEAAMNNQVFTFTWREDIRPQDPTKKPCDA 526  
DB 140 GTGLCADTKHGALGSPRLRLGECVGRGEAAMNNQVFTFTWREDIRPQDPTKKPCDA 199  
QY 527 ISHTSPVTLVYDCHSMKGNQLMKTKRKDKTLVHPVSGSCMDCSSESHRIFMNTCNPSLTQ 586  
DB 200 ISHTSPVTLVYDCHSMKGNQLMKTKRKDKTLVHPVSGSCMDCSSESHRIFMNTCNPSLTQ 259  
QY 587 WLFETHNSTVLEKFNRN 603  
DB 260 WLFETHNSTVLEKFNRN 276

RESULT 14

ABG08189 standard; protein, 335 AA.

ABG08189;

13-FEB-2002 (first entry)

Novel human diagnostic protein #8180.

Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KM food supplement; medical imaging; diagnostic; genetic disorder.

Homo sapiens.

OS



QY 338 GLBIWGEQYEIFKV 353  
 Db 241 GLBIWGEQYEIFKV 256

## RESULT 16

ID ABB66873 standard; protein; 666 AA.

AC ABB66873;

DT 26-MAR-2002 (first entry)

DE Drosophila melanogaster polypeptide SEQ ID NO 27411.

XX Drosophila; developmental biology; cell signalling; insecticide;

KW pharmaceutical.

OS Drosophila melanogaster.

PN WO200171042-A2.

PD 27-SEP-2001.

PF 23-MAR-2001; 2001WO-US009231.

PR 23-MAR-2000; 2000US-0191637P.

PR 11-JUL-2000; 2000US-00614150.

XX (PEKE ) PE CORP NY.

PI Venter JC, Adams M, Li PWD, Myers EW;

XX WPI; 2001-656860/75.

DR N-PSDB; ABL10976.

PT New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from Drosophila and for elucidating cell signalling and cell-cell  
 PT interactions.

PS Disclosure; SEQ ID NO 27411; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from Drosophila. The invention is  
 CC useful in developmental biology and in elucidating cell signalling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
 CC sequences (ABL01840-ABL16175) and the encoded proteins (ABBS7737-  
 CC ABB72072). The sequence data for this patent did not form part of the  
 CC printed specification, but was obtained in electronic format directly  
 CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences

XX Sequence 666 AA;

Query Match 41.5%; Score 1360; DB 4; Length 666;

Best Local Similarity 49.0%; Pred. No. 3.3e-129;

Matches 277; Conservative 76; Mismatches 160; Indels 52; Gaps 15;

QY 67 DGQKLDKMDHDEKAIIRDAQVNGEGRPYPMTDAERVDQVR--ENGFNITYSDKISL 123

Db 121 DASVAKDMDHDTFMEKDAKRVGLGEGKASTLDDESQRDLERKMSLNGFPALSDSISV 180

QY 124 NRSI.PDIIRHPCNSKRYLETLPNTSIIIPFNEGWSLLRTVSHVSLNSPELVAEIVLV 183

Db 181 NRSVPDIDHPLCRKKEVYAKLPYVSVIIIFINEYISVLMRSVSHSLINSPELMKEITLV 240

QY 184 DDFSDREHLKKTPLBEDYMA-LFPYSVILRTKTKGGLIRTRMGASVATGDTVITFLDSHCEA 242

Db 241 DDHSRREVLGKELRYIAEHFKWVAVLPRRTGLIGARAGARAYATAEVLIFLDSHVEA 300

QY 243 NVNMLPRLLDRIANRRKTIYVCPMIDVIVHDDFRYTGQGDAMRGAFDMEWTYKRIPIRPE 302

Db 301 NVNMLPRLLEPIALNKRITAVCPFIDVIDHTNFHYRAQ-DEGARGAFMEFFYKRLPLLP 359

QY 303 LQKADPSDPSPSPVMAAGLPAVDKRMFWELGYPDGLIWMGGEQYEIFKVMCGRMED 362

Db 360 DLK-HPADPFKSPIMAGGLPAISHEFFWELGIDEGDIDWGEQYEIFKIMWGGMED 418

QY 363 IPCGRVGHYR-----KRYEYKVPAGVSLANLKRVAEVMWDEYAEYIYQRPE-YHLS 416

Db 419 APCSRIGHIYRGRPHQPSPRK---GDYLHNKYKRVAVEMWDEYKNLYSHGDLYESVD 475

QY 417 AGDVAVQKLRSSLNCSFPMFKTIAMDLPKPYRPPAPAAAMGELRNVT-GLCADT- 474

Db 476 PGDLTEQKAIKTKLNCSEFFWEEVAFDLMKTYPPVDPSPYAMGALQNGNONLCLDTL 535

QY 475 ---KGALG-----SPLRLEGCVRGGRGAAMNNQVFTFRREDIRPGDQTKKF 522

Db 536 GRKGNMGMTACADNKTQR-----TQFWELSMKDLR---LRKKE 576

QY 523 CFDA-ISHTSPVTLYDCHSNKGNQIMKY-RKDKTYHPVSG-SCMDSSDRIRFMNTC 578

Db 577 CLDVQIWDANAPVWLMDCHSGGQGVYDYDRHKLQKGTGRCLELTPFSQGEVANKC 636

QY 579 NPSSLTOQMLFEHTNSIVLEKFNKN 603

Db 637 DTNRFQQMNFGSFNKTDLDNYSOD 661

## RESULT 17

ID ABB5867 standard; protein; 666 AA.

AC ABB5867;

DT 26-MAR-2002 (first entry)

DE Drosophila melanogaster polypeptide SEQ ID NO 3393.

XX Drosophila; developmental biology; cell signalling; insecticide;

KW pharmaceutical.

OS Drosophila melanogaster.

PN WO200171042-A2.

PD 27-SEP-2001.

PF 23-MAR-2001; 2001WO-US009231.

PR 23-MAR-2000; 2000US-0191637P.

PR 11-JUL-2000; 2000US-00614150.

XX (PEKE ) PE CORP NY.

PI Venter JC, Adams M, Li PWD, Myers EW;

XX WPI; 2001-656860/75.

DR N-PSDB; ABL02970.

PT New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from Drosophila and for elucidating cell signalling and cell-cell  
 PT interactions.

PS Disclosure; SEQ ID NO 3393; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from Drosophila. The invention is  
 CC useful in developmental biology and in elucidating cell signalling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
 CC sequences (ABL01840-ABL16175) and the encoded proteins (ABBS7737-  
 CC ABB72072). The sequence data for this patent did not form part of the  
 CC printed specification, but was obtained in electronic format directly

CC from WIFO at ftp.wifo.int/pub/published\_pct\_sequences  
 XX Sequence 666 AA:

Query Match 41.5%; Score 1360; DB 4; Length 666;  
 Best Local Similarity 49.0%; Pred. No. 3.3e-129;  
 Matches 277; Conservative 76; Mismatches 160; Indels 52; Gaps 15;

QY 67 DGQGLKMDHKEARRDAQRYNGEGRPYMTAERVDAYR--ENGFFIYSDKISL 123  
 DB 121 DASVKKMDHYTTEFKDAKRVGLGKASTLDESGRDLEKRSLSLNGFALLSDISIV 180  
 QY 124 NRSPLDRIHPNCNSKRYLETLPNTSIIIPFNHGMSSILRTVSHVLSRSPPELVAEIYV 183  
 DB 181 NRSVPDIRHPICRCKEYVAKLPYTSVLIIFNYEILSLMSVSHLSNRPPELKEIILV 240  
 QY 184 DDFSDREHLKKPLEDYNA-LPPSVRIILRTKKREGILRTMLGASVATGDTVITFLDSHCEA 242  
 DB 241 DSHSDREYLGKELEIYIAEHKMWVRVRLPRRTGLIGRAAGANATAVAVLIFLDSHYEA 300  
 QY 243 NNVNWLPLDLRIANRKTIVCEMIDVIDHDDFRYETQAGDAMGAFPMWYTKRIPIIPE 302  
 DB 301 NNVNWLPLDLRIANRKTIVCEMIDVIDHDDFRYETQAGDAMGAFPMWYTKRIPIIPE 359  
 QY 303 LQKADPSDPSPSPVMAAGLFAVDRKFMELGSDYDGLIENGSEGYEISFKYMGCGHMD 362  
 DB 360 DLK-HPADPFSKSPIMAGGLFAISREFFMELGSDYDGLIENGSEGYEISFKYMGCGHMD 418  
 QY 363 IPCSRVGHYR-----KYVPYKPVAGVSLAANLKRVAEVMWDEYAEIYQRRPE-YRHLS 416  
 DB 419 APCSRGHYIRGRPNHQPSPRK---GDYLNKRYKRVAEVMWDEYKNTLYSHGDLYESVD 475  
 QY 417 AGDYAVVQKLRSSLANCSFKFMETKIAMDLPEKYPVPEPPAANGIRNNGT-GLCADT- 474  
 DB 476 PGDLTEKALRTKLNCKSKFMFEEVAFDLMKTYPPVDPSPYAMGALQNGNQMLCLDTL 535  
 QY 475 ---KHGALG-----SPLRLEGCVRGGEAAMNNQVPTFTWRDIPRGDQHTKKF 522  
 DB 536 GRKKHKKMGAYACADNITKTPOR-----TQFMELSKKDLR---LRRKKE 576  
 QY 523 CFDA--ISHTSPVTLVDCHSKMGKQLWKY-RKDKTLVHPVSG-SCMDCSESDHRIFNATC 578  
 DB 577 CLVDQVIMDANAPVWLMPCGSGGQNGVWYDYRHKQLKHGTGRCCLBELPSPSGEVANKC 636  
 QY 579 NPSGLTQOMLFHTNSTVLEKENEN 603  
 DB 637 DTNRFQOMNFGSFNKTALDNYSD 661

## RESULT 18

AAU07774  
 ID AAU07774 standard; protein; 321 AA.

AAU07774;

DT 04-DEC-2001 (first entry)

XX Human novel transferase protein, NHP #17.

XX Human; transferase; breast cancer; prostate cancer; immunogen;

XX Gene therapy; antisense.

XX Homo sapiens.

XX WO200164903-A2.

XX 07-SEP-2001.

XX 28-FEB-2001; 2001WO-US006460.

XX 29-FEB-2000; 2000US-0185920P.

XX 02-MAR-2000; 2000US-0186558P.

XX 24-MAR-2000; 2000US-0191849P.

XX (LEXI-) LEXICON GENETICS INC.

XX Donoho G, Hilbun E, Turner CA, Friedrich G, Abuin A;

XX Zambrowicz B, Sands AT, Walke DW, Wilganowski NL, Hu Y, Kieke JA;

XX Potter DG;

XX WPI; 2001-550185/61.

XX N-PSDB; AAS12621.

XX Novel nucleic acid sequences encoding novel human proteins useful for

XX diagnosis, drug screening, clinical trial monitoring and treatment of

XX diseases and disorders.

XX Disclosures; Page 50-51; 60pp; English.

XX The invention relates to isolated nucleic acids encoding novel human

XX transferase proteins (NHP). The nucleic acids and proteins are useful for

XX diagnosis, drug screening, clinical trial monitoring and treatment of

XX diseases and disorders e.g. breast and prostate cancer. NHPs can also be

XX useful for augmenting the efficacy of chemotherapeutic agents used in

XX treatment of breast or prostate cancer. The nucleic acid is also useful

XX in NHP gene regulation, and as antisense primers in amplification

XX reactions of NHP gene sequences. NHPs are useful for producing

XX antibodies. The present sequence represents a novel human transferase

XX which has sequence similarity to N-acetyl-galactosaminyltransferase

XX

XX Sequence 321 AA;

XX Query Match 40.0%; Score 1311; DB 4; Length 321;

XX Best Local Similarity 99.2%; Pred. No. 1e-124;

XX Matches 251; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 MRRERKRLQAVLVLAVALVLPVGMALYRERQDPGTGGSGAVALPAGGSHSRQK 60

DB 1 MRRERKRLQAVLVLAVALVLPVGMALYRERQDPGTGGSGAVALPAGGSHSRQK 60

QY 61 KTFPLGSGQGLKMDHKEARRDAQRYNGEGRPYMTAERVDAYRNGFFIYVSDK 120

DB 61 KTFPLGSGQGLKMDHKEARRDAQRYNGEGRPYMTAERVDAYRNGFFIYVSDK 120

QY 121 ISLNRSLPDIRHPNCNSKRYLETLPNTSIIIPFNHGMSSILRTVSHVLSRSPPELVAEI 180

DB 121 ISLNRSLPDIRHPNCNSKRYLETLPNTSIIIPFNHGMSSILRTVSHVLSRSPPELVAEI 180

QY 181 VLVDFFSDREHLKKPLEDYNALPPSVRIILRTKKREGILRTMLGASVATGDTVITFLDSHC 240

DB 181 VLVDFFSDREHLKKPLEDYNALPPSVRIILRTKKREGILRTMLGASVATGDTVITFLDSHC 240

QY 241 EAVNWLPLDLR 253

DB 241 EAVNWLPLDLR 253

XX 27-SEP-2001.

## RESULT 19

ABB59363  
 ID ABB59363 standard; protein; 1252 AA.

ABB59363;

DT 26-MAR-2002 (first entry)

XX Drosophila melanogaster polypeptide SEQ ID NO 4881.

XX Drosophila; developmental biology; cell signalling; insecticide;

XX pharmaceutical.

XX Drosophila melanogaster.

XX WO200171042-A2.

XX 27-SEP-2001.



QY 175 ELVAEIVLVDFSGREHLKKPLEDYMLPFSVRLTRKKRGLIRTRMLGASVATGVIT 234  
 Db 21 EMGOEITKLKDGSGREHLKKPLEDYMLPFSVRLTRKKRGLIRTRMLGASVATGVIT 80  
 QY 235 FLDSHCANVWMLPPLLDRIARNKRTIVCPMIDVIDHDDPRYETQADAMKGFDMWEMY 294  
 Db 81 FLDSHCANVWMLPPLLDRIARNKRTIVCPMIDVIDHDDPRYETQADAMKGFDMWEMY 140  
 QY 295 KRPIPELOKADPSDPPE-----SPMAGGLF 322  
 Db 141 KRPIPELOKADPSDPPECOLVRLTGLDADSQQHDSVSHLRCQADFLFPMSPVMAGGLF 200  
 QY 323 AVBRKMWELGMY-----DP----- 337  
 Db 201 AVBRKMWELGMYFSGHDDPQPKESLFFPSSCTSPQVESHMPAGMPPIPEQYCNMSPA 260  
 QY 338 -----GLEIMG-----EYVE-- 348  
 Db 261 VTGCVTAGVYCLGWMKXKXKMGKTDLLGARWKSXKFXPTNQDVKNPHDLIEGQPK 320  
 QY 349 ---ISPRVW-----CGRMEDIPC 365  
 Db 321 EHCFTESAWLGCPCPONGSMLQSORSHREBNHGIEMTVISAVSCGLHGVCGGMEDIPC 380  
 QY 366 SRVGHYRKTVYKVPAGVSLARLKRVAEVMDEVAEYIYQRRPEYRHLGAGVAVQKK 425  
 Db 381 S-----RNLKRVAEVMDEVAEYIYQRRPEYRHLGAGVAVQKK 419  
 QY 426 LRSSLNCKSPKFWMTKIAMDLPKFYPEVPEPAAAMGEIRNVGTGLCADTKHAGLSPLRL 485  
 Db 420 LRSSLNCKSPKFWMTKIAMDLPKFYPEVPEPAAAMGEIRNVGTGLCADTKHAGLSPLRL 476  
 QY 486 EGCY-----RGR 492  
 Db 477 GSCTHHWGRGR 488

## RESULT 21

ID ADP65304 standard; protein; 559 AA.  
 AC ADP65304;

XX 12-AUG-2004 (first entry)  
 DE Human polypeptide N-acetylgalactosaminyltransferase 1.

XX autoimmune disease; arthritis; gene expression analysis;  
 KM rheumatoid arthritis; collagen-induced; immunosuppressive; antirheumatic;  
 KM antirheumatic; osteoarthritis; anti-gout; antiinflammatory; dermatological;  
 KM immunomodulatory; lupus; ankylosing spondylitis; fibrositis;  
 KM fibromyalgia; osteoarthritis; gout; juvenile rheumatoid arthritis;  
 KM immune; human.

OS Homo sapiens.  
 PN W02003072827-A1.

XX 04-SEP-2003.  
 PD 31-OCT-2002; 2002WC-US035433.

XX 31-OCT-2001; 2001US-0336220P.  
 PR (CHIL-) CHILDREN'S HOSPITAL MEDICAL CENT.

XX Hirsch R. Thorton SL;  
 XX WPI, 2003-712740/67.  
 DR GENBANK; NP\_065207.

XX Diagnosing and analyzing autoimmune disease using gene expression  
 PT profiles and microarray technology, useful for diagnosing and treating

PT rheumatoid arthritis, lupus, fibrositis, osteoarthritis, fibromyalgia and  
 PT gout.

XX Disclosure; Page; 56pp; English.

XX The invention relates to a novel method for diagnosing and analysing  
 CC autoimmune disease or arthritides. The method comprises obtaining a  
 CC patient sample containing mRNA, analysing gene expression using the mRNA  
 CC that results in a gene expression signature to diagnose or analyse the autoimmune disease  
 CC or arthritides in the patient, where gene expression of at least 60% of  
 CC the genes correlates with that of the gene signature. The invention  
 CC further comprises: a treatment of rheumatoid arthritis; identification of  
 CC genes for targeting in the treatment of rheumatoid arthritis in a mammal  
 CC other than a mouse; diagnosis of rheumatoid arthritis in a mammal; an  
 CC array or gene chip, specific for rheumatoid arthritis; diagnosis or  
 CC analyses of autoimmune disease or rheumatoid arthritis; screening the  
 CC efficacy of a candidate drug in vitro for the treatment of collagen-  
 CC induced arthritis; and reducing the symptoms associated with collagen-  
 CC induced arthritis. The compositions of the invention have the following  
 CC activities: immunosuppressive, antirheumatic, antirheumatic, osteopathic,  
 CC antigout, antiinflammatory, dermatological, and immunomodulatory. The  
 CC methods and compositions of the present invention are useful for  
 CC diagnosing and treating autoimmune disease or arthritides, such as  
 CC rheumatoid arthritis, lupus, ankylosing spondylitis, fibrositis,  
 CC fibromyalgia, osteoarthritis, gout, juvenile rheumatoid arthritis, and an  
 CC immune disease caused by an infectious agent. This sequence represents a  
 CC protein sequence relating to the genes used in the analysis and treatment  
 CC of autoimmune diseases or arthritides. Note: This sequence is not shown  
 CC in the specification. It has been supplied in an electronic format from  
 CC WIPO.

XX Sequence 559 AA;

Query Match 34.3%; Score 1125; DB 7; Length 559;  
 Best Local Similarity 44.0%; Pred. No. 2,8e-105;  
 Matches 232; Conservative 87; Mismatches 170; Indels 38; Gaps 16;

QY 88 GNGEQRP--YPMDAERVDQAVRENGFNIVSDKISLNSLPDIRHPCNSKRYLETLP 145  
 Db 57 GPGEMGRKPVVPRKDEKMKEMFKINQFNLMASEMILANSLDPVRLGCKTKVYPPNLP 116  
 QY 146 NTSIIIPHNHGSILIRTVSHVNSPPELVAEIVLVDFSGREHLKKLEDMV--ALFP 204  
 Db 117 TTSVIVVFNHNAWSTLIRTVSHVNSPPELVAEIVLVDFSGREHLKKLEDMV--ALFP 176  
 QY 205 SVRLIRTKRREGLIRTRMLGASVATGVITFLDSHCANVWMLPPLLDRIARNKRTIVCP 264  
 Db 177 PAVIRNEQSGILIRTRMLGASVATGVITFLDSHCANVWMLPPLLDRIARNKRTIVCP 236  
 QY 265 MIDVIDHDDPRYETQAG-DAMKGFDMWEMYKRIPIPP--ELQKADSPDPSSPVNAGG 320  
 Db 237 IIDVIDSDTPEY--MAGSDWTYGFNWKLFNRYVPYQREMRRKGRDTPVPTPWAGG 294  
 QY 321 LFAVDRKMWELGMYFSGHDDPQPKESLFFPSSCTSPQVESHMPAGMPPIPEQYCNMSPA 380  
 Db 295 LFAVDRKMWELGMYFSGHDDPQPKESLFFPSSCTSPQVESHMPAGMPPIPEQYCNMSPA 354  
 QY 381 PAGVS--LARNLRVAEVMDEVAEYIYQRRPEYRHLGAGVAVQKKLRSSLNCKSPKWF 438  
 Db 355 PGGTGQIINKNRRRLAEVWMDERKFNPIYISPEVTAKVDYDISRVLGRKLOCKKPSWY 414  
 QY 439 MTKIAMD--LPKFPVPEPAAAMGEIRNVGTGLCADTKHAGLSPLRLGCVGRGEAA 496  
 Db 415 LEMIVPDSQIPRYH-----FSLGEIRNVETNQCLDNMARKENKEKGIENC-HGMG--- 463  
 QY 497 WNNKQVPTFTWRDRIKRGDQHTKRCFPAISITSPVTLVDCSMKGNQIAWKRYKOK--TL 555  
 Db 464 --GNQVFSYANKERIRDD-----LCLDVSKLNGVPTMLKCHRLKGNQIMBEYDPVYTL 515  
 QY 556 YHPVSGSCMD-CESGDHRI-FNATCNPSLSLTOOMLFHTSTVLEK 600  
 Db 516 QHYNNOCLDKATBEDSQVPSINDCN-GRSQOMLAR--NVTLPBIF 559

RESULT 22

AAB43561 standard; protein; 561 AA.

AC AAB43561;

DE 08-FEB-2001 (first entry)

Human cancer associated protein sequence SEQ ID NO:1006.

Human; cancer associated gene; cancer antigen; detection; cancer; diagnosis; cytotoxic; proliferative; antineoplastic; immunomodulator; antidiabetic; antiaesthetic; antirheumatic; anticholesteric; antiviral; antiinflammatory; antithyroid; antiallergic; antibacterial; cardiac; dermatologic; neuroprotective; thrombolytic; coagulant; nociceptive; vasotonic; antipsoriatic; angiogenic; gene therapy; inflammation; immune disorder; hematopoietic cell disorder; autoimmune disorder; allergic reaction; graft versus host disease; organ rejection; haemostatic; thrombolytic; cardiovascular disorder; infection; neurologic disease; drug screening.

OS Homo sapiens.

W0200055350-A1.

21-SEP-2000.

08-MAR-2000; 2000WO-US005882.

12-MAR-1999; 99US-0124270P.

(HUMA-) HUMAN GENOME SCI INC.

Rosen CA, Ruben SM;

WPI; 2000-587533/55.

N-PSDB; AAC77770.

Novel isolated nucleic acids comprising sequences encoding peptides useful for treating or diagnosing e.g. cancer.

Claim 11; Page 1581-1584; 23522P; English.

AAC77607 to AAC78448 encode the human cancer associated proteins given in AAB43398 to AAB44239. The proteins can have activities based on the tissues and cells the genes are expressed in. Example of activities include: cytotoxic; proliferative; antineoplastic; anticholesteric; antidiabetic; antiaesthetic; antirheumatic; anticholesteric; antiinflammatory; antithyroid; antiallergic; antibacterial; antiviral; dermatologic; neuroprotective; cardiac; thrombolytic; coagulant; nociceptive; vasotonic; antipsoriatic and angiogenic. The polynucleotides and polypeptides can be used for preventing, treating or ameliorating medical conditions and diagnosing pathological conditions. Polynucleotides, polypeptides, antibodies, agonists and antagonists from the present invention may be used to treat immune disorders by activating or inhibiting the proliferation, differentiation or mobilization of immune cells to treat disorders of hematopoietic cells, autoimmune disorders, allergic reactions, graft versus host disease and organ rejection, modulate haemostatic or thrombolytic activity, modulate inflammation, cancers, cardiovascular disorders, neurological disease and bacterial or viral infections. The peptides, nucleotides, antibodies, agonists and antagonists may be also be used in drug screens. AAC78449 to AAC78457 and AAB44240 represent sequences used in the exemplification of the present invention

Sequence 561 AA;

Query Match 34.3%; Score 1125; DB 3; Length 561;

Best Local Similarity 44.0%; Pred. No. 2.8e-105; Indels 38; Gaps 16; Matches 232; Conservative 87; Mismatches 170;

QY 88 GNGEGRP--YPTDAERVDQAYRENGFNIVSDKISLNSLPDIRHPNCSKRYLETLP 145  
 DB 59 GPGEMKRPVVI PKEDGKMKEMKFNQFNIMASEMIALNLSLPDVRLEGCKTQYVPPNLP 118  
 QY 146 NTSIIIPFNHEGSSLLRTYTHSVLNSPPPELVAEIVYVDFSPREHKKLEDEM-ALFP 204  
 DB 119 TTSVIVFHNEMWSTLRTYTHSVLNSPPPELVAEIVYVDFSPREHKKLEDEM-ALFP 178  
 QY 205 SVRLRTKREGLRTMTLGSVATGDPVITFLDSHCANVNWLPPLDIRIARNTIYVCP 264  
 DB 179 PNVHIREGSGILIRALKGAASVKGQVITFLDAHCECTYGMLEPLARIKHRRITVCP 238  
 QY 265 MIDVIDHDFRYETQAG-DAMRGAFWEMTYKAIPIFP---ELQKADSPDPFSPVWAG 320  
 DB 239 IIDVISDTEFY--MAGSDMTYGGFMKLNFRWYPPVQREMDRKGDRTPVPTVAGG 296  
 QY 321 LFANDRKPFELGVDGLIHWGEOYEISPKVWMSGRMEDIPCSRVGHYRKYVPYKV 380  
 DB 297 LFSIDRDYFQELIGYDGMIDWGENLEISFRIMQCGTLEIVTCHVGHVFRKATPYTF 356  
 QY 381 PAGVS--LAENLKRVAEVMDEYAEYIYQRPPEYRHLISAGDVAVQKLRSLNCKSPKP 438  
 DB 357 PGGTGQIINNKNRRLAEVWDEFPNFFYIISPQTKVDYGDISSRVGLRHKLOCKPESWY 416  
 QY 439 MTKIAMD-LPKYPPVEPPAAMGEIRNVGTGLCADTKGALGSPRLRGCVRGGEAA 496  
 DB 417 LENIYPSQIPRHY-----FSLGEIRNVETNOCIDNMARKENKGYIFNC-HGMK--- 465  
 QY 497 WNNQVFTFWRREDIRPQDQHTKKPCFPAISHTSPVTLVDCSHMKGNOLMKRYKDK-TL 555  
 DB 466 --GNQVFSYTNKKEIRTD-----LCLDVSCKLNGPVTMLKCHLNQMLEIDPVKLT 517  
 QY 556 YHPVSGSCMD-CSESDHRI-FMNTCNPSLTOQWLPENTNSTYLEKP 600  
 DB 518 QHVNNOCLDKATEBDSQVPSINDCN-GSRSGQWLAR-NVTLPEIF 561

RESULT 23

AAB79836 standard; protein; 559 AA.

AC AAB79836;

DE 04-DEC-2003 (first entry)

Rat polypeptide GalNAc transferase T1, SEQ ID 76.

Analgesic; pain; streptozocin-induced diabetes; rat.

Rattus norvegicus.

EP1279744-A2.

29-JAN-2003.

26-JUN-2002; 2002EP-00255249.

27-JUN-2001; 2001GB-00018354.

07-FEB-2002; 2002GB-00002910.

(WARN) WARNER LAMBERT CO.

Brookbank RA, Dixon AK, Lee K, Pinnock RD;

WPI; 2003-395407/38.

N-PSDB; AAB79837.

Use of isolated gene sequences and encoded polypeptides that are upregulated in the spinal cord in response to streptozocin-induced diabetes for screening compounds for the treatment of pain, or for diagnosing pain.

Claim 1; Page 148-149; 334pp; English.

XX The present invention relates to nucleotide sequences which are useful in  
 CC the screening of compounds for the treatment of pain, or for the  
 CC diagnosis of pain. The nucleotide sequences are up-regulated in the  
 CC spinal cord in response to streptozocin-induced diabetes. The present  
 CC sequence was used to illustrate the invention.

XX Sequence 559 AA;

Query Match 34.0%; Score 1116; DB 7; Length 559;  
 Best Local Similarity 43.6%; Pred. No. 2,3e-104;  
 Matches 230; Conservative 88; Mismatches 171; Indels 38; Gaps 16;

```

QY GNGEGRP--YPMTDARVDQAYRENGFNIVSDKISLNSLPDIRHPNCKSKRYLETLP 145
DB 57 GPGEMGRPVVLPKEDQKMKEMFKINQFNLMASMIAFNRSLLPVRLGCKTKYVPPSLP 116
QY 146 NTSLIFPNHNGSSLLRTVSHVNLNRSPPPELVAEIVLVDPSPDRHKKPLLEDYM-ALFP 204
DB 117 TTSVIVFHNHAWSTLLRTVSHVNLNRSPPPELVAEIVLVDPSPDRHKKPLLEDYM 176
QY 205 SVRLRTKKEGGLIRTMGLASVATGVTITFLDSHCANVMPLPDLRIARNKTIYCP 264
DB 177 PNHVIRMEGRSGILIRALTKGAASVKGQVITFLDAHCECTGWLPLARIKHDRRTVCP 236
QY 265 MIDVIDHDDFRYETQAG-DAMRGAFWEMYYKRIPIPP--ELQKADSPDFESPVMAG 320
DB 237 IIDVISDITFEY--MAGSDMTYGGFNWKLNFWRVYPVQREMDRRKGRDRTLPRVPTWAG 294
QY 321 LFAVDRKFWELGSDYDGLIENGEOYEISFKVMCGRMEDIPCSRVGHYRYKYVYKV 380
DB 295 LFSIDRYFOEIGTYDAGMDIWGEMLEISFRIWCCGTLEIYCSHVGHVFRATPTTF 354
QY 381 PAGVS--LAENLKRVAEVMDEVAEYIYORPREYRHLASGVAVQKLRSLNCKSPKMF 438
DB 355 PGGTGOIINKNNRRLAEVWMDERKNFYIISPGVTKYVDYDISRVLGRHKLQCKPFSWY 414
QY 439 MTKIAMD--LPKFPVPEPPAAAMGEIRNVGTGICADTKHAGLSPRLBGCVRGEBA 496
DB 415 LENIYPPDSQIPIRHY-----FSLGEIRNVETNQCCLNMARKENKKGIFNCG-HGMG--- 463
QY 497 MNMNQVFTFWREDIRPGDPQHTKKCFDAISHTSPVTLVDCSHMKGNOLMKYAKDK-TL 555
DB 464 --GNQVFSYANKKIRITD-----LCIDVSKLNGPVTMLKCHLKNQMLEYDPAVLT 515
QY 556 YHPVSGCMD-CSESDHRI-FMNTCNPSLSLTOQWLFEHTNSTVLEKF 600
DB 516 QHVNNSQCCLDKATEBDSQVPSIRDCT-GSRSQQLLR--NVTLPETIF 559

```

RESULT 24

ADB79796 standard; protein; 559 AA.

ADB79796;

04-DEC-2003 (first entry)

Rat polypeptide GalNAC transferase T1, SEQ ID 36.

Analgesic; pain; streptozocin-induced diabetes; rat.

Rattus norvegicus.

EP1279744-A2.

29-JAN-2003.

26-JUL-2002; 2002EP-00255249.

27-JUL-2001; 2001GB-00018354.

07-FEB-2002; 2002GB-00002910.

PA (WARN) WARNER LAMBERT CO.

XX Brooksbank RA, Dixon AK, Lee K, Pinnock RD;

XX WPI; 2003-395407/38.

DR N-PSDB; ADB79797.

XX Use of isolated gene sequences and encoded polypeptides that are  
 PT upregulated in the spinal cord in response to streptozocin-induced  
 PT diabetes for screening compounds for the treatment of pain, or for  
 PT diagnosing pain.

PS Claim 1; Page 91-93; 334pp; English.

XX The present invention relates to nucleotide sequences which are useful in  
 CC the screening of compounds for the treatment of pain, or for the  
 CC diagnosis of pain. The nucleotide sequences are up-regulated in the  
 CC spinal cord in response to streptozocin-induced diabetes. The present  
 CC sequence was used to illustrate the invention.

XX Sequence 559 AA;

Query Match 34.0%; Score 1116; DB 7; Length 559;  
 Best Local Similarity 43.6%; Pred. No. 2,3e-104;  
 Matches 230; Conservative 88; Mismatches 171; Indels 38; Gaps 16;

```

QY GNGEGRP--YPMTDARVDQAYRENGFNIVSDKISLNSLPDIRHPNCKSKRYLETLP 145
DB 57 GPGEMGRPVVLPKEDQKMKEMFKINQFNLMASMIAFNRSLLPVRLGCKTKYVPPSLP 116
QY 146 NTSLIFPNHNGSSLLRTVSHVNLNRSPPPELVAEIVLVDPSPDRHKKPLLEDYM-ALFP 204
DB 117 TTSVIVFHNHAWSTLLRTVSHVNLNRSPPPELVAEIVLVDPSPDRHKKPLLEDYM 176
QY 205 SVRLRTKKEGGLIRTMGLASVATGVTITFLDSHCANVMPLPDLRIARNKTIYCP 264
DB 177 PNHVIRMEGRSGILIRALTKGAASVKGQVITFLDAHCECTGWLPLARIKHDRRTVCP 236
QY 265 MIDVIDHDDFRYETQAG-DAMRGAFWEMYYKRIPIPP--ELQKADSPDFESPVMAG 320
DB 237 IIDVISDITFEY--MAGSDMTYGGFNWKLNFWRVYPVQREMDRRKGRDRTLPRVPTWAG 294
QY 321 LFAVDRKFWELGSDYDGLIENGEOYEISFKVMCGRMEDIPCSRVGHYRYKYVYKV 380
DB 295 LFSIDRYFOEIGTYDAGMDIWGEMLEISFRIWCCGTLEIYCSHVGHVFRATPTTF 354
QY 381 PAGVS--LAENLKRVAEVMDEVAEYIYORPREYRHLASGVAVQKLRSLNCKSPKMF 438
DB 355 PGGTGOIINKNNRRLAEVWMDERKNFYIISPGVTKYVDYDISRVLGRHKLQCKPFSWY 414
QY 439 MTKIAMD--LPKFPVPEPPAAAMGEIRNVGTGICADTKHAGLSPRLBGCVRGEBA 496
DB 415 LENIYPPDSQIPIRHY-----FSLGEIRNVETNQCCLNMARKENKKGIFNCG-HGMG--- 463
QY 497 MNMNQVFTFWREDIRPGDPQHTKKCFDAISHTSPVTLVDCSHMKGNOLMKYAKDK-TL 555
DB 464 --GNQVFSYANKKIRITD-----LCIDVSKLNGPVTMLKCHLKNQMLEYDPAVLT 515
QY 556 YHPVSGCMD-CSESDHRI-FMNTCNPSLSLTOQWLFEHTNSTVLEKF 600
DB 516 QHVNNSQCCLDKATEBDSQVPSIRDCT-GSRSQQLLR--NVTLPETIF 559

```

RESULT 25

ADB62526 standard; protein; 559 AA.

ADB62526;

29-JAN-2004 (first entry)

Rat Protein Q10473, SEQ ID NO 8456.



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 22, 2004, 13:39:57 ; Search time 21 Seconds  
(without alignments)  
1904.274 Million cell updates/sec

Title: US-10-001-851-2

Perfect score: 3278  
Sequence: 1 MRKKEKRLQAVLVLAALV.....TQGMLEHTNSTYLEKERN 603

Scoring table: BIOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 478139 seqs, 66318000 residues

Total number of hits satisfying chosen parameters: 478139

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 100 summaries

Database :

Issued Patents AA:\*  
1: /cgn2\_6/ptodata/1/iaa/5A\_COMB.pep:\*  
2: /cgn2\_6/ptodata/1/iaa/5B\_COMB.pep:\*  
3: /cgn2\_6/ptodata/1/iaa/6A\_COMB.pep:\*  
4: /cgn2\_6/ptodata/1/iaa/6B\_COMB.pep:\*  
5: /cgn2\_6/ptodata/1/iaa/PTCUS\_COMB.pep:\*  
6: /cgn2\_6/ptodata/1/iaa/backfile1.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3278	100.0	603	4	US-09-795-926-43
2	3242.5	98.9	631	4	US-09-795-926-41
3	2771	84.5	506	4	US-09-795-926-31
4	2746.5	83.8	535	4	US-09-795-926-29
5	1877	57.3	366	4	US-09-795-926-39
6	1370	41.8	269	4	US-09-795-926-27
7	1311	40.0	321	4	US-09-795-926-35
8	1115	34.0	517	2	US-08-967-508-19
9	1115	34.0	517	3	US-08-967-506-19
10	1115	34.0	517	5	PCT-US94-02552-19
11	1115	34.0	559	2	US-08-967-508-9
12	1115	34.0	559	3	US-08-967-506-9
13	1115	34.0	559	5	PCT-US94-02552-9
14	1026	31.3	578	4	US-09-217-3068-2
15	1024	31.2	560	4	US-09-217-3068-10
16	990	30.2	240	4	US-09-795-926-33
17	988	30.1	209	4	US-09-795-926-37
18	971	29.6	552	4	US-10-140-002-196
19	941	28.7	302	4	US-09-270-767-33733
20	926	28.2	639	3	US-09-376-856-2
21	920	28.1	638	3	US-09-347-488-2
22	916	27.9	633	2	US-08-648-298-2
23	804	24.5	224	4	US-09-795-926-23
24	744	22.7	603	4	US-09-973-457-2
25	675	20.6	289	4	US-09-270-767-45334
26	561	17.1	238	4	US-09-270-767-46411
27	503.5	15.4	366	4	US-09-270-767-43543

28	483	14.7	143	4	US-09-795-926-21	Sequence 21, Appl
29	481	14.7	142	4	US-09-795-926-25	Sequence 25, Appl
30	450.5	13.7	119	4	US-09-973-457-6	Sequence 6, Appl
31	428.5	13.1	161	4	US-09-270-767-58910	Sequence 58910, A
32	287	8.8	122	4	US-09-513-999C-7570	Sequence 7570, Ap
33	277.5	8.5	148	4	US-09-270-767-32872	Sequence 32872, A
34	241.5	7.4	313	4	US-09-270-767-46209	Sequence 46209, A
35	234.5	7.2	300	4	US-09-270-767-43809	Sequence 43809, A
36	228.5	7.0	290	4	US-09-270-767-61773	Sequence 61773, A
37	218.5	6.7	72	4	US-09-270-767-33350	Sequence 61989, A
38	168	5.1	73	4	US-09-270-767-48567	Sequence 33350, A
39	167	5.1	73	4	US-09-270-767-33441	Sequence 48567, A
40	167	5.1	101	4	US-09-270-767-48568	Sequence 33441, A
41	167	5.1	101	4	US-09-134-000C-5947	Sequence 48568, A
42	133	4.1	424	4	US-09-270-767-48567	Sequence 5947, Ap
43	129	3.9	88	4	US-09-270-767-33601	Sequence 33601, A
44	129	3.9	88	4	US-09-270-767-48818	Sequence 48818, A
45	128	3.9	135	4	US-09-973-457-5	Sequence 5, Appl
46	127	3.9	436	4	US-09-338-352-4416	Sequence 4416, Ap
47	125.5	3.8	965	4	US-09-437-277-3	Sequence 3, Appl
48	124	3.8	702	4	US-09-437-277-1	Sequence 1, Appl
49	123.5	3.8	534	4	US-09-134-000C-5087	Sequence 5087, Ap
50	123	3.8	721	4	US-09-107-532A-6889	Sequence 6889, Ap
51	121.5	3.7	1056	4	US-09-134-000C-5086	Sequence 5086, Ap
52	119.5	3.6	492	2	US-08-468-812-4	Sequence 4, Appl
53	119.5	3.6	492	2	US-08-468-812-7	Sequence 7, Appl
54	119.5	3.6	492	2	US-08-590-563-4	Sequence 4, Appl
55	119.5	3.6	492	3	US-08-590-563-7	Sequence 4, Appl
56	119.5	3.6	492	4	US-09-770-621-4	Sequence 4, Appl
57	119.5	3.6	492	4	US-09-770-621-7	Sequence 7, Appl
58	119.5	3.6	492	4	US-09-235-832-4	Sequence 4, Appl
59	119.5	3.6	492	4	US-09-235-832-7	Sequence 7, Appl
60	118	3.6	480	2	US-08-468-812-5	Sequence 5, Appl
61	118	3.6	480	3	US-08-590-563-5	Sequence 5, Appl
62	118	3.6	480	4	US-09-770-621-5	Sequence 5, Appl
63	118	3.6	480	4	US-09-235-832-5	Sequence 5, Appl
64	116.5	3.6	88	4	US-09-270-767-33890	Sequence 33890, A
65	116.5	3.6	88	4	US-09-270-767-49107	Sequence 49107, A
66	112.5	3.4	909	3	US-09-425-383-2	Sequence 2, Appl
67	111.5	3.4	491	2	US-08-468-812-8	Sequence 8, Appl
68	111.5	3.4	491	3	US-08-590-563-8	Sequence 8, Appl
69	111.5	3.4	491	4	US-09-770-621-8	Sequence 8, Appl
70	111.5	3.4	491	4	US-09-235-832-8	Sequence 8, Appl
71	107.5	3.3	507	4	US-09-130-337A-25	Sequence 25, Appl
72	107	3.3	270	4	US-09-495-406-25	Sequence 25, Appl
73	107	3.3	270	4	US-09-816-028A-39	Sequence 39, Appl
74	107	3.3	270	4	US-10-303-162-39	Sequence 39, Appl
75	105.5	3.2	187	4	US-09-973-457-4	Sequence 4, Appl
76	105.5	3.2	336	4	US-09-543-681A-6535	Sequence 6535, Ap
77	105	3.2	127	1	US-08-392-828C-39	Sequence 39, Appl
78	105	3.2	127	3	US-09-330-945-39	Sequence 39, Appl
79	105	3.2	324	1	US-08-597-236-10	Sequence 10, Appl
80	105	3.2	324	1	US-08-746-682A-10	Sequence 10, Appl
81	102.5	3.1	540	1	US-08-378-761A-77	Sequence 77, Appl
82	102.5	3.1	540	1	US-08-485-286-77	Sequence 77, Appl
83	102	3.1	317	4	US-09-583-110-4849	Sequence 4849, Ap
84	101.5	3.1	395	4	US-08-635-552A-4	Sequence 4, Appl
85	101.5	3.1	419	2	US-08-270-581-2	Sequence 2, Appl
86	101.5	3.1	419	4	US-09-146-893-2	Sequence 2, Appl
87	101.5	3.1	419	4	US-08-675-499A-5	Sequence 5, Appl
88	101.5	3.1	419	4	US-08-812-008-5	Sequence 5, Appl
89	101	3.1	328	4	US-09-583-110-3554	Sequence 3554, Ap
90	100	3.1	331	4	US-09-634-238-747	Sequence 247, Ap
91	99.5	3.0	303	4	US-09-495-406-17	Sequence 17, Appl
92	99.5	3.0	303	4	US-09-816-028A-29	Sequence 29, Appl
93	99.5	3.0	303	4	US-10-303-162-29	Sequence 29, Appl
94	99	3.0	341	4	US-09-328-352-1178	Sequence 7178, Ap
95	99	3.0	301	4	US-09-495-406-15	Sequence 15, Appl
96	98	3.0	301	4	US-09-816-028A-27	Sequence 27, Appl
97	98	3.0	301	4	US-10-303-162-27	Sequence 27, Appl
98	96.5	2.9	623	1	US-08-653-740-7	Sequence 7, Appl
99	96.5	2.9	623	2	US-09-073-594-7	Sequence 7, Appl
100	96.5	2.9	623	3	US-09-275-925-7	Sequence 7, Appl

## ALIGNMENTS

## RESULT 1

US-09-795-926-43  
Sequence 43, Application US/09795926  
Patent No. 6555669  
GENERAL INFORMATION:  
APPLICANT: Donoho, Gregory  
APPLICANT: Hilbun, Erin  
APPLICANT: Turner, C. Alexander Jr.  
APPLICANT: Friedrich, Glenn  
APPLICANT: Abuhin, Alejandro  
APPLICANT: Zambrowicz, Brian  
APPLICANT: Sands, Arthur T.  
APPLICANT: Walke, D. Wade  
APPLICANT: Milganowski, Nathaniel L.  
APPLICANT: Hu, Yi  
APPLICANT: Kiele, James Alvin  
APPLICANT: Potter, David George  
TITLE OF INVENTION: NOVEL HUMAN TRANSFERASE PROTEINS AND  
FILE REFERENCE: POLYNUCLEOTIDES ENCODING THE SAME  
CURRENT FILING DATE: 2001-02-28  
CURRENT APPLICATION NUMBER: US/09/795,926  
PRIOR FILING DATE: 2000-02-29  
PRIOR APPLICATION NUMBER: US 60/185,920  
PRIOR FILING DATE: 2000-03-02  
PRIOR APPLICATION NUMBER: US 60/191,849  
PRIOR FILING DATE: 2000-03-24  
NUMBER OF SEQ ID NOS: 47  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 43  
LENGTH: 603  
TYPE: PRT  
ORGANISM: homo sapiens  
US-09-795-926-43

Query Match 100.0%; Score 3278; DB 4; Length 603;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 603; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MRRREKLLQAVLVLAALVLLPVGGLMALYRERQPDGPGSGAAVAAPAAQGSRSRQK 60  
DB 1 MRRREKRLQAVLVLAALVLLPVGGLMALYRERQPDGPGSGAAVAAPAAQGSRSRQK 60  
QY 61 KTFELGDSQKLDKMDHDEKAIIRDAQRVNGEGGRPYPMTDARVDQAYRENGFNIIYVSDK 120  
DB 61 KTFELGDSQKLDKMDHDEKAIIRDAQRVNGEGGRPYPMTDARVDQAYRENGFNIIYVSDK 120  
QY 121 ISLRSLSPLDIHHPNCNSKRYLETLPNTSIIIPFNHSGSSLLRTVHSLVNSPPELVAEI 180  
DB 121 ISLRSLSPLDIHHPNCNSKRYLETLPNTSIIIPFNHSGSSLLRTVHSLVNSPPELVAEI 180  
QY 181 VLVDPSDRREHLKKPLLEDYMLPFSVRLIRTKKREGILRTMLGASVATGVITFLDSHC 240  
DB 181 VLVDPSDRREHLKKPLLEDYMLPFSVRLIRTKKREGILRTMLGASVATGVITFLDSHC 240  
QY 241 EANNVMLPPLLDRIARNRKTIIVCPMIDVIDHDDFRYETQAGDAMRGAFDWMYYKRIPIP 300  
DB 241 EANNVMLPPLLDRIARNRKTIIVCPMIDVIDHDDFRYETQAGDAMRGAFDWMYYKRIPIP 300  
QY 301 PELQKADSDPFESPVMAAGLFAVDRAKMFELGGYDPLGLIWMGEGQYEISFKVMMCGGRM 360  
DB 301 PELQKADSDPFESPVMAAGLFAVDRAKMFELGGYDPLGLIWMGEGQYEISFKVMMCGGRM 360  
QY 361 EDIPCSRGHTYRKVPYKVPAGVSLANTLKRVAEVMWDEYAEIYYQRRPPTRHLSADV 420  
DB 361 EDIPCSRGHTYRKVPYKVPAGVSLANTLKRVAEVMWDEYAEIYYQRRPPTRHLSADV 420

QY 421 AVQKKLRSLNCKSFKFMFKTKIAMDLPKFPVPEPPAAANGELINNVGTGLCADTKHGALG 480  
DB 421 AVQKKLRSLNCKSFKFMFKTKIAMDLPKFPVPEPPAAANGELINNVGTGLCADTKHGALG 480  
QY 481 SPLRLGECVGRGGAANNNOVFPTTWREDIRPDGPDHTKKFCFDALSHSPVTLVYCHS 540  
DB 481 SPLRLGECVGRGGAANNNOVFPTTWREDIRPDGPDHTKKFCFDALSHSPVTLVYCHS 540  
QY 541 MKGNQLMKYRKDKTLVHPVSGSCMDGSESDHRIIPMTCNPSLSLTQQLFEHTNSTVLEKF 600  
DB 541 MKGNQLMKYRKDKTLVHPVSGSCMDGSESDHRIIPMTCNPSLSLTQQLFEHTNSTVLEKF 600  
QY 601 NRN 603  
DB 601 NRN 603

## RESULT 2

US-09-795-926-41  
Sequence 41, Application US/09795926  
Patent No. 6555669  
GENERAL INFORMATION:  
APPLICANT: Donoho, Gregory  
APPLICANT: Hilbun, Erin  
APPLICANT: Turner, C. Alexander Jr.  
APPLICANT: Friedrich, Glenn  
APPLICANT: Abuhin, Alejandro  
APPLICANT: Zambrowicz, Brian  
APPLICANT: Sands, Arthur T.  
APPLICANT: Walke, D. Wade  
APPLICANT: Milganowski, Nathaniel L.  
APPLICANT: Hu, Yi  
APPLICANT: Kiele, James Alvin  
APPLICANT: Potter, David George  
TITLE OF INVENTION: NOVEL HUMAN TRANSFERASE PROTEINS AND  
FILE REFERENCE: POLYNUCLEOTIDES ENCODING THE SAME  
CURRENT FILING DATE: 2001-02-28  
CURRENT APPLICATION NUMBER: US/09/795,926  
PRIOR FILING DATE: 2000-02-29  
PRIOR APPLICATION NUMBER: US 60/185,920  
PRIOR FILING DATE: 2000-03-02  
PRIOR APPLICATION NUMBER: US 60/191,849  
PRIOR FILING DATE: 2000-03-24  
NUMBER OF SEQ ID NOS: 47  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 41  
LENGTH: 631  
TYPE: PRT  
ORGANISM: homo sapiens  
US-09-795-926-41

Query Match 98.9%; Score 3242.5; DB 4; Length 631;  
Best Local Similarity 95.4%; Pred. No. 0;  
Matches 601; Conservative 0; Mismatches 29; Indels 1;  
QY 1 MRRREKLLQAVLVLAALVLLPVGGLMALYRERQPDGPGSGAAVAAPAAQGSRSRQK 60  
DB 1 MRRREKRLQAVLVLAALVLLPVGGLMALYRERQPDGPGSGAAVAAPAAQGSRSRQK 60  
QY 61 KTFELGDSQKLDKMDHDEKAIIRDAQRVNGEGGRPYPMTDARVDQAYRENGFNIIYVSDK 120  
DB 61 KTFELGDSQKLDKMDHDEKAIIRDAQRVNGEGGRPYPMTDARVDQAYRENGFNIIYVSDK 120  
QY 121 ISLRSLSPLDIHHPNCNSKRYLETLPNTSIIIPFNHSGSSLLRTVHSLVNSPPELVAEI 180  
DB 121 ISLRSLSPLDIHHPNCNSKRYLETLPNTSIIIPFNHSGSSLLRTVHSLVNSPPELVAEI 180  
QY 181 VLVDPSDRREHLKKPLLEDYMLPFSVRLIRTKKREGILRTMLGASVATGVITFLDSHC 240  
DB 181 VLVDPSDRREHLKKPLLEDYMLPFSVRLIRTKKREGILRTMLGASVATGVITFLDSHC 240

```

QY 241 BANNVMLPELDRIRANRKTIVCEPMIDVIDHDHOFRETQAGDMRGAPEMDEMYKKRPIP 3000
Db 241 BANNVMLPELDRIRANRKTIVCEPMIDVIDHDHOFRETQAGDMRGAPEMDEMYKKRPIP 3000
QY 301 PELQADPSPDPESPVMAGGLFAVDKRKFWELGSGYPGLIEWGEGOEYISFK----- 3522
Db 301 PELQADPSPDPESPVMAGGLFAVDKRKFWELGSGYDPLGLEIWGEGOEYISFKGLHMLPRL 3606
QY 353 -----YVMCGGRMEDIPCSRVGHIYRKTYVPYKVPAGVSLARNLK 3911
Db 361 VSNMSPQAVFLPRAPNMLAQVMCCGRMBEDIPCSRVGHIYRKTYVPYKVPAGVSLARNLK 4200
QY 392 RVAEYVMDEYAEYIYORREYRHLISAGDVAVOKKLSLNLCKSFKFMPTKIANDLPRFYP 4511
Db 421 RVAEYVMDEYAEYIYORREYRHLISAGDVAVOKKLSLNLCKSFKFMPTKIANDLPRFYP 4800
QY 452 PVEPRAAAMGEIRINVTGTCADTKHIALGSLPLREBGCVGRGGAANNMNVFFFTWREDI 5111
Db 481 PVEPRAAAMGEIRINVTGTCADTKHIALGSLPLREBGCVGRGGAANNMNVFFFTWREDI 5406
QY 512 RPDGDQHTKKCFPDALSHTSPTVLYDCHSMKGNQMLKTRKDKTLVHPVSSGCMDCSESDH 5711
Db 541 RPDGDQHTKKCFPDALSHTSPTVLYDCHSMKGNQMLKTRKDKTLVHPVSSGCMDCSESDH 6000
QY 572 RIFNMTCNPSSLTQOQLFEHTNSTVLEKTN 601
Db 601 RIFNMTCNPSSLTQOQLFEHTNSTVLEKTN 630

```

RESULT 3  
US-09-795-926-31

```

: GENERAL INFORMATION:
: APPLICANT: Donoho, Gregory
: APPLICANT: Hilbun, Erin
: APPLICANT: Turner, C. Alexander Jr.
: APPLICANT: Friedrich, Glenn
: APPLICANT: Abuhin, Alejandro
: APPLICANT: Zambrowicz, Brian
: APPLICANT: Sands, Arthur T.
: APPLICANT: Walke, D. Wade
: APPLICANT: Milganowski, Nathaniel L.
: APPLICANT: Hu, Yi
: APPLICANT: Kiehe, James Alvin
: APPLICANT: Potter, David George
: TITLE OF INVENTION: NOVEL HUMAN TRANSFERASE PROTEINS AND
: FILE OF INVENTION: POLYNUCLEOTIDES ENCODING THE SAME
: FILE REFERENCE: LEX-0144-USA
: CURRENT APPLICATION NUMBER: US/09/795,926
: CURRENT FILING DATE: 2001-02-28
: PRIOR APPLICATION NUMBER: US 60/185,920
: PRIOR FILING DATE: 2000-02-29
: PRIOR APPLICATION NUMBER: US 60/186,558
: PRIOR FILING DATE: 2000-03-02
: PRIOR APPLICATION NUMBER: US 60/191,849
: PRIOR FILING DATE: 2000-03-24
: NUMBER OF SEQ ID NOS: 47
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 31
: LENGTH: 506
: TYPE: RPT
: ORGANISM: homo sapiens
US-09-795-926-31

```

Query Match	84.5%;	Score 2771;	DB 4;	length 506;
Best Local Similarity	100.0%;	Pred. No. 2.7e-289;		
Matches 506;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

QY 98 MTDARVDQAYRENGFNIIYVDGKSLNRSLLPIRHFNQNSKYLETTPNTSTIIIPFNHEG 157  
Dbb 1 MTDARVDQAYRENGFNIIYVDGKSLNRSLLPIRHFNQNSKYLETTPNTSTIIIPFNHEG 60

QY	158	WSSLIRTAHVSLNRSPEPLVAEIVLVDDPSDEHLKKPLEDMVALPFSVRILIRTKREGI	217
Db	61	WSSLIRTAHVSLNRSPEPLVAEIVLVDDPSDEHLKKPLEDMVALPFSVRILIRTKREGI	120
QY	218	IRTRMIGASVATGDIVTIFFLDSHCENANWMLPELLDRIANRRTIVCPMIDVIDHDDFRYE	277
Db	121	IRTRMIGASVATGDIVTIFFLDSHCENANWMLPELLDRIANRRTIVCPMIDVIDHDDFRYE	180
QY	278	TOAGAMGADPWEMTYRIRIPRPELOKADBEDPESPMMAGLFAVDKMTWELGGYBP	337
Db	181	TOAGAMGADPWEMTYRIRIPRPELOKADBEDPESPMMAGLFAVDKMTWELGGYBP	240
QY	338	GLEIWGSGQYELISFKVMWCGGMEBIPCSRUGHIRKTYPYKVPAGVSLARNLKRAVEYW	397
Db	241	GLEIWGSGQYELISFKVMWCGGMEBIPCSRUGHIRKTYPYKVPAGVSLARNLKRAVEYW	300
QY	398	MDEVAEYIYORRPEYRHLISAGDVAVOKLRSLSLCKSFKFEMFTKIAMDLPKYPPYEPRA	457
Db	301	MDEVAEYIYORRPEYRHLISAGDVAVOKLRSLSLCKSFKFEMFTKIAMDLPKYPPYEPRA	360
QY	458	AAMGBIRNVRGSLCADTRHGALGSLPLREGCYGRGGEAANNNMOYFTFWREDIRGGDPQ	517
Db	361	AAMGBIRNVRGSLCADTRHGALGSLPLREGCYGRGGEAANNNMOYFTFWREDIRGGDPQ	420
QY	518	HTKKCFDAISHTSPVLYDCHSMKGNOLMKYRKDKTLYHPVSGSCMDCSESDDHRIFNMT	577
Db	421	HTKKCFDAISHTSPVLYDCHSMKGNOLMKYRKDKTLYHPVSGSCMDCSESDDHRIFNMT	480
QY	578	CNPSLLTOOMLPEHTNSVYLEKFNPN	603
Db	481	CNPSLLTOOMLPEHTNSVYLEKFNPN	506

RESULT 4  
US-09-795-926-29

```

Sequence 29, Application US/09795926
Patent No. 6555669
GENERAL INFORMATION:
APPLICANT: Donoho, Gregory
APPLICANT: Hilbun, Erin
APPLICANT: Turner, C. Alexander Jr.
APPLICANT: Friedrich, Glenn
APPLICANT: Abuhin, Alejandro
APPLICANT: Zambrowicz, Brian
APPLICANT: Sands, Arthur T.
APPLICANT: Walke, D. Wade
APPLICANT: wilganowski, Nathaniel L.
APPLICANT: Hu, Yi
APPLICANT: Kieke, James Alvin
APPLICANT: Potter, David George
TITLE OF INVENTION: NOVEL HUMAN TRANSFERASE PROTEINS AND
TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING THE SAME
FILE REFERENCE: LEX-0144-USA
CURRENT APPLICATION NUMBER: US/09/795,926
CURRENT FILING DATE: 2001-02-28
PRIOR APPLICATION NUMBER: US 60/185,920
PRIOR FILING DATE: 2000-02-29
PRIOR APPLICATION NUMBER: US 60/186,558
PRIOR FILING DATE: 2000-03-02
PRIOR APPLICATION NUMBER: US 60/191,849
PRIOR FILING DATE: 2000-03-24
NUMBER OF SEQ ID NOS: 47
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 29
LENGTH: 535
TYPE: PRT
ORGANISM: homo sapiens
US-09-795-926-29

```

Query Match	83.8%	Score 2746.5;	DB 4;	Length 535;
Best Local Similarity	94.6%;	Pred. No. 1.3e-286;		
Matches 506; Conservative	0;	Mismatches 0;	Indels 29;	Gaps 1;

```

Qy 98 MTDARVDQAYRENGFNIIYVSDKISLNRSLPDIRHPNCNSKRYLETLPNTSIIIPFNEG 157
Db 1 MTDARVDQAYRENGFNIIYVSDKISLNRSLPDIRHPNCNSKRYLETLPNTSIIIPFNEG 60
Qy 158 WSSILRTVHSHVLSNPPELVAEIVLVDPSDREHLKKPLEDYMALPFSVRILRTKKEGL 217
Db 61 WSSILRTVHSHVLSNPPELVAEIVLVDPSDREHLKKPLEDYMALPFSVRILRTKKEGL 120
Qy 218 IRTMGLASVATGVTITFLDSCANVWMLPILLDIRARNKTIYCPMIDVIDHDDPRYE 277
Db 121 IRTMGLASVATGVTITFLDSCANVWMLPILLDIRARNKTIYCPMIDVIDHDDPRYE 180
Qy 278 TOAGDARGAFDWMYKRIPIPELOKADSPDFESFVMAAGLFAVDRKMFELGGYDP 337
Db 181 TOAGDARGAFDWMYKRIPIPELOKADSPDFESFVMAAGLFAVDRKMFELGGYDP 240
Qy 338 GLEIWGBOYEISFK-----VMCGMEDI PCSRV 368
Db 241 GLEIWGBOYEISFKGLHMLPRLVSNWPQAVPLPRAPNMLALQVMCGMEDI PCSRV 300
Qy 369 GHIRKXVVPYKPGVSLARLKVAAEYVMDVAYEYIQRRPEYRHLSAGDAVOKLRS 428
Db 301 GHIRKXVVPYKPGVSLARLKVAAEYVMDVAYEYIQRRPEYRHLSAGDAVOKLRS 360
Qy 429 SLNCKSFKFWTKIAMDLPKFYPPVEPPAAMGEIRNVGTGLCADTKHGLSGPLRLEG 488
Db 361 SLNCKSFKFWTKIAMDLPKFYPPVEPPAAMGEIRNVGTGLCADTKHGLSGPLRLEG 420
Qy 489 VRGGEAAMNNQVFTFWREDIRPGDPOHTKCFEADISHTSPVTLTDCHSMKGNQIMK 548
Db 421 VRGGEAAMNNQVFTFWREDIRPGDPOHTKCFEADISHTSPVTLTDCHSMKGNQIMK 480
Qy 549 YRKOKTLHPVSGSCMDCSDHRIFMNTCNPSLTOOMLEHNTSTYLEKFNEN 603
Db 481 YRKOKTLHPVSGSCMDCSDHRIFMNTCNPSLTOOMLEHNTSTYLEKFNEN 535

RESULT 5
US-09-795-926-39
; Sequence 39, Application US/09795926
; Patent No. 655569
; GENERAL INFORMATION:
; APPLICANT: Donoho, Gregory
; APPLICANT: Hilbun, Erin
; APPLICANT: Turner, C. Alexander Jr.
; APPLICANT: Friedrich, Glenn
; APPLICANT: Abuin, Alejandro
; APPLICANT: Zambrowicz, Brian
; APPLICANT: Sands, Arthur T.
; APPLICANT: Walke, D. Wade
; APPLICANT: Milganowski, Nathaniel L.
; APPLICANT: Hu, Yi
; APPLICANT: Kieke, James Alvin
; APPLICANT: Potter, David George
; TITLE OF INVENTION: NOVEL HUMAN TRANSFERASE PROTEINS AND
; FILE REFERENCE: LEX-0144-USA
; CURRENT APPLICATION NUMBER: US/09/795,926
; CURRENT FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: US 60/185,920
; PRIOR FILING DATE: 2000-02-29
; PRIOR APPLICATION NUMBER: US 60/186,558
; PRIOR FILING DATE: 2000-03-02
; PRIOR APPLICATION NUMBER: US 60/191,849
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 47
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 39
; LENGTH: 366
; TYPE: PRT
; ORGANISM: homo sapiens
US-09-795-926-39

```

```

Query Match 57.3%; Score 1877; DB 4; Length 366;
Best Local Similarity 100.0%; Pred. No. 3, 5e-193;
Matches 353; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MRRKREKLLQAVLVALVLLPVGGLMALYRERQDPGTGGSGAIVAPAGGSHSRQK 60
Db 1 MRRKREKLLQAVLVALVLLPVGGLMALYRERQDPGTGGSGAIVAPAGGSHSRQK 60
Qy 61 KTFPLGGQKDKMDHKEAIRDAQRVNGEQQRPYMTAERYDQAYRENGFNIIYVSDK 120
Db 61 KTFPLGGQKDKMDHKEAIRDAQRVNGEQQRPYMTAERYDQAYRENGFNIIYVSDK 120
Qy 121 ISLNRSLPDIRHPNCNSKRYLETLPNTSIIIPFNEGSSILRTVHSHVLSNPPELVAEI 180
Db 121 ISLNRSLPDIRHPNCNSKRYLETLPNTSIIIPFNEGSSILRTVHSHVLSNPPELVAEI 180
Qy 181 VLVDPSDREHLKKPLEDYMALPFSVRILRTKKEGLIRTMGLASVATGVTITFLDSC 240
Db 181 VLVDPSDREHLKKPLEDYMALPFSVRILRTKKEGLIRTMGLASVATGVTITFLDSC 240
Qy 241 EANNVNLPLLDRIARNRKTIYCPMIDVIDHDDPRYETOAGDARGAFDWMYKRIPI 300
Db 241 EANNVNLPLLDRIARNRKTIYCPMIDVIDHDDPRYETOAGDARGAFDWMYKRIPI 300
Qy 301 PELQKADSPDFESFVMAAGLFAVDRKMFELGGYDGLIEWGBOYEISFKV 353
Db 301 PELQKADSPDFESFVMAAGLFAVDRKMFELGGYDGLIEWGBOYEISFKV 353

```

```

RESULT 6
US-09-795-926-27
; Sequence 27, Application US/09795926
; Patent No. 655569
; GENERAL INFORMATION:
; APPLICANT: Donoho, Gregory
; APPLICANT: Hilbun, Erin
; APPLICANT: Turner, C. Alexander Jr.
; APPLICANT: Friedrich, Glenn
; APPLICANT: Abuin, Alejandro
; APPLICANT: Zambrowicz, Brian
; APPLICANT: Sands, Arthur T.
; APPLICANT: Walke, D. Wade
; APPLICANT: Milganowski, Nathaniel L.
; APPLICANT: Hu, Yi
; APPLICANT: Kieke, James Alvin
; APPLICANT: Potter, David George
; TITLE OF INVENTION: NOVEL HUMAN TRANSFERASE PROTEINS AND
; FILE REFERENCE: LEX-0144-USA
; CURRENT APPLICATION NUMBER: US/09/795,926
; CURRENT FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: US 60/185,920
; PRIOR FILING DATE: 2000-02-29
; PRIOR APPLICATION NUMBER: US 60/186,558
; PRIOR FILING DATE: 2000-03-02
; PRIOR APPLICATION NUMBER: US 60/191,849
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 47
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 27
; LENGTH: 269
; TYPE: PRT
; ORGANISM: homo sapiens
US-09-795-926-27

```

```

Query Match 41.8%; Score 1370; DB 4; Length 269;
Best Local Similarity 100.0%; Pred. No. 9, 1e-139;
Matches 256; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 98 MTDARVDQAYRENGFNIIYVSDKISLNRSLPDIRHPNCNSKRYLETLPNTSIIIPFNEG 157
Db 1 MTDARVDQAYRENGFNIIYVSDKISLNRSLPDIRHPNCNSKRYLETLPNTSIIIPFNEG 60

```

OY		158	WSSLRLRTVHSVLRNRPPELVAETIVVDPSDEHLKKPLIEDYMALPSPRIILRTKREG.	21
Db		61	WSSLRLRTVHSVLRNRPPELVAETIVVDPSDEHLKKPLIEDYMALPSPRIILRTKREG.	120
OY		218	IRTRMLGASVATGDVITFLDSCICANVMNLPELLDIRIARNRKTIVCPMIDVIDHDPRYE	277
Db		121	IRTRMLGASVATGDVITFLDSCICANVMNLPELLDIRIARNRKTIVCPMIDVIDHDPRYE	180
OY		278	TQAGAMGAPFPMWNTYRKRIPIPIPELOKADPSDPESPVMAGGLPAVDKRWELGGTYP	337
Db		181	TQAGAMGAPFPMWNTYRKRIPIPIPELOKADPSDPESPVMAGGLPAVDKRWELGGTYP	240
OY		338	GLEIWGEQGYEISIFKY 353	
Db		241	GLEIWGEQGYEISIFKY 256	
 RESULT 7 US-09-795-926-35				
		/ Sequence 35, Application US/09795926		
		/ Patent No. 6555669		
		/ GENERAL INFORMATION:		
		/ APPLICANT: Donoho, Gregory		
		/ APPLICANT: Hilburn, Erin		
		/ APPLICANT: Turner, C. Alexander Jr.		
		/ APPLICANT: Friederich, Glenn		
		/ APPLICANT: Abuhin, Alejandro		
		/ APPLICANT: Zambrowicz, Brian		
		/ APPLICANT: Sands, Arthur T.		
		/ APPLICANT: Walke, D. Wade		
		/ APPLICANT: Milganowski, Nathaniel L.		
		/ APPLICANT: Hu, Yi		
		/ APPLICANT: Kieker, James Alvin		
		/ APPLICANT: Potter, David George		
		/ TITLE OF INVENTION: NOVEL HUMAN TRANSFERASE PROTEINS AND		
		/ FILE REFERENCE: POLYNUCLEOTIDES ENCODING THE SAME		
		/ CURRENT APPLICATION NUMBER: US/09/795, 926		
		/ PRIOR FILING DATE: 2001-02-28		
		/ PRIOR APPLICATION NUMBER: US 60/185,920		
		/ PRIOR FILING DATE: 2000-02-29		
		/ PRIOR APPLICATION NUMBER: US 60/186,558		
		/ PRIOR FILING DATE: 2000-03-02		
		/ PRIOR APPLICATION NUMBER: US 60/191,849		
		/ PRIOR FILING DATE: 2000-03-24		
		/ NUMBER OF SEQ ID NOS: 47		
		/ SOFTWARE: FastSeq for Windows Version 4.0		
		/ SEQ ID NO 35		
		/ LENGTH: 321		
		/ TYPE: PRY		
		/ ORGANISM: homo sapiens		
		US-09-795-926-35		
 Query Match            40.0%; Score 1311; DB 4; Length 321; Best Local Similarity   99.2%; Pred. No. 2.8e-132; Matches   251; Conservative   1; Mismatches   1; Indels   0; Gaps   0				
OY		1	MRRKRKLQVALVTALAIVLPNVGMALYHEROPDTGGSSGAAVAPAOGSGHSROX	60
Db		1	MRRKRKLQVALVTALAIVLPNVGMALYHEROPDTGGSSGAAVAPAOGSGHSROX	60
OY		61	KTFPLGDDQCKLDWMHDKAIRDAORVGNGEGCRPYPMTDARVDOAYENGFNITYSDX	120
Db		61	KTFPLGDDQCKLDWMHDKAIRDAORVGNGEGCRPYPMTDARVDOAYENGFNITYSDX	120
OY		121	ISLNRSPLDIRHPNCNSKRYLETLPNTSIIIPFNHGWSLLRTVHSVLRNRPPELVAEI	180
Db		121	ISLNRSPLDIRHPNCNSKRYLETLPNTSIIIPFNHGWSLLRTVHSVLRNRPPELVAEI	180
OY		181	VLVNDPSREHKRLLEDYMALFPESVRILTKRGGLIRTRMLGASVATGDVITFLDSC	240
Db		181	VLVNDPSREHKRLLEDYMALFPESVRILTKRGGLIRTRMLGASVATGDVITFLDSC	240

```

QY      241 EANNWMLPPLDRL 253      ||||||| :
Db      241 EANNWMLPPLDGL 253

```

---

```

RESULT 8
US-08-967-508-19
Sequence 19, Application US/08967508
Patent No. 5910570
GENERAL INFORMATION:
APPLICANT: The Upjohn Company
APPLICANT: For U.S. PURPOSES ONLY: Elhammer, Ake P. and Homa, Fred L.
TITLE OF INVENTION: A Cloned DNA Encoding a UDP-galNAc:
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSER: Pharmacia & Upjohn Company, Intellectual
STREET: 301 Henrietta Street
CITY: Kalamazoo
STATE: Michigan
COUNTRY: USA
ZIP: 49001
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/967,508
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/602,830
FILING DATE: 13 No. 5910570ember 1995
ATTORNEY/AGENT INFORMATION:
NAME: Darnley Jr., James D.
REGISTRATION NUMBER: 33,673
REFERENCE/DOCKET NUMBER: 4755.P CP
TELECOMMUNICATION INFORMATION:
TELEPHONE: 616-833-2210
TELEFAX: 616-833-8897
TELEX: 224401
INFORMATION FOR SEQ ID NO: 19:
SEQUENCE CHARACTERISTICS:
LENGTH: 517 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-967-508-19

```

---

```

Query Match      34.0%; Score 1115; DB 2; Length 517;
Best Local Similarity 43.6%; Pred. No. 8.3e-11;
Matches 230; Conservative 89; Mismatches 170; Indels 38; Gaps 16

```

---

```

QY      88 GNGEGRPP--YPMIDARVDQAYRENGFNIVSDKISLNSLPDIRHPNCNSKRYLETLP 145
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      15 GPGMGKRPVIVPKDQKMKEMFKINGFNIMASIMIALNSLPVRLRGCTKYYPDLRP 74
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY      146 NTSIIIPHNHNGSSLIKRTVHSAVNSRPPPELVAVIIVLVDPSDREHLKKPLEDTM-ALFP 204
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      75 TTSVVIYFHNHNAWSTLLKRTVHSAVINRSPRHMLEIIVLVDASERDPLKRPLESYVKKLKV 134
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY      205 SVRLIRKKRGGILRTMTLQASVATGVTITFLDSHCANWMLPPLDIRIRNRKKTIVCP 264
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      135 PVAHVRHQRSGLIKRRLKGAASVKGQVITFLDAHCECTGWLPEPLARIHDKRTIVCP 194
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY      265 MIDVLDHDDFERYIQAG-DAMRGAFDWEMYYKRIPIPP---ELQKADPSDFESPVAVAG 320
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      195 IIVDISDITPEY--MAGSDMTYGGFNKMLNPRWYFVQREMDRDKGRGRTLPVRIPTMAGG 252
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY      321 LPAVDRKFMELGQYDPLGLEIWSGEQYELISFVMMCGRMEDIPCSRGHVITRYKTVPVKV 380
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

Db 253 LFSIDRDFGEIGYDAGMDIMGGENLEISFRIMQCGGTLEIVTCSHVGHFRKATPYTF 312  
Qy 381 PAGVS--LARNLKVAEVMDEVAEYIYQRRPEYRHLSAGDVAVQKLRSSLNCKSPKWF 438  
Db 313 FGGTGQILNKNRRRLAEVWMDPEFNFFYIISPVTKYDYGDISLGRHLKQCRPFSWY 372  
Qy 439 MTKIAMD--LPKFPVPPEPAAWGEIRNVTGTCADTKHGALGSLRLBGCVRGGEAA 496  
Db 373 LENIYPSQIPRHY-----FSLGEIRNVTNOCIDNMARKENKVGIFNC-HGMG--- 421  
Qy 497 WNNQVFFFTWRBEDIREDPDQHTKKFCFDAISHTSPVTLVYDCHSMKGNQMLKRYKDK-TL 555  
Db 422 --GNQVFSYTNKKEIRTD-----LCIDVSKLNGPVTMLKCHLKGQMLBEPVVLTL 473  
Qy 556 YHVPVSGCMD-CSESDHRI-FMNTCNPSLSLQOQLFHTNSTVLEKF 600  
Db 474 QHVNNOCLDKATBEDSQVPSIRDCS-GSRSQOMLLR--NVTLPBIF 517

## RESULT 9

US-08-967-506-19  
Sequence 19, Application US/08967506  
Patent No. 6096512  
GENERAL INFORMATION:  
APPLICANT: The Upjohn Company  
APPLICANT: FOR U.S. PURPOSES ONLY: Elhammer, Ake P. and Homa, Fred L.  
TITLE OF INVENTION: A cloned DNA Encoding a UDP-galNAC:  
TITLE OF INVENTION: Polypeptide, N-Acetylglactosaminyltransferase  
NUMBER OF SEQUENCES: 19  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Pharmacia & Upjohn Company, Intellectual  
ADDRESSEE: Property Legal Services  
STREET: 301 Henrietta Street  
CITY: Kalamazoo  
STATE: Michigan  
COUNTRY: USA  
ZIP: 49001  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/967,506  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/602,830  
FILING DATE: 13 No. 6096512ember 1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Darnley Jr., James D.  
REGISTRATION NUMBER: 33,673  
REFERENCE/DOCKET NUMBER: 4755.P CP  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 616-833-2210  
TELEFAX: 616-833-8897  
TELEX: 224401  
INFORMATION FOR SEQ ID NO: 19:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 517 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-967-506-19

Query Match 34.0%; Score 1115; DB 3; Length 517;  
Best Local Similarity 43.6%; Pred. No. 8,3e-111;  
Matches 230; Conservative 89; Mismatches 170; Indels 38; Gaps 16;  
Qy 88 GNGEGRP--YPMDDAEVDAVYRENGFNIVYSDKISLNRSLPIDIRHNCNSKRYLETLP 145

Db 15 GPGEMGRPVVLPKEDQKMKEMKINQFNLMASEMIALNRLSPDVRLEGCKTKYVDPNLP 74  
Qy 146 NTSIIIPFHEMGSSILRTVHVSILNRPPELVAEIVLVDDPSREHKKPLEDDYM-ALFP 204  
Db 75 TTSVIVIFNHEANSTLTARTYHVSINRSPRMLBEIVLVDDASBDFLAKRPLESVKRLKV 134  
Qy 205 SVRIIRTKKBBGLIRITMLASVATGVTITFLDSHCANVNLPLLDRIARNKTIYCP 264  
Db 135 PVHVRMEORSGILRAIKGAASVSKQVITFLDAHCECTGWMLEPLARIKHDRKTVCP 194  
Qy 265 MIDVIDHDDFRYERQAG-DAMRGAPEWEMYYKRIPIPE---ELQKADSPDPSSVWAG 320  
Db 195 IIVDISDTEY--MASSDMTYGSEFNKMLNRPVYQREMDRKGDRITLPRTVWAG 252  
Qy 321 LFAVDRKFMWELAGYDPEGLBIEWGEQYEISFKVMCCGRMEDIPCSRVGHIYRKVYKYV 380  
Db 253 LFSIDRDFGEIGYDAGMDIMGGENLEISFRIMQCGGTLEIVTCSHVGHFRKATPYTF 312  
Qy 381 PAGVS--LARNLKVAEVMDEVAEYIYQRRPEYRHLSAGDVAVQKLRSSLNCKSPKWF 438  
Db 313 FGGTGQILNKNRRRLAEVWMDPEFNFFYIISPVTKYDYGDISLGRHLKQCRPFSWY 372  
Qy 439 MTKIAMD--LPKFPVPPEPAAWGEIRNVTGTCADTKHGALGSLRLBGCVRGGEAA 496  
Db 373 LENIYPSQIPRHY-----FSLGEIRNVTNOCIDNMARKENKVGIFNC-HGMG--- 421  
Qy 497 WNNQVFFFTWRBEDIREDPDQHTKKFCFDAISHTSPVTLVYDCHSMKGNQMLKRYKDK-TL 555  
Db 422 --GNQVFSYTNKKEIRTD-----LCIDVSKLNGPVTMLKCHLKGQMLBEPVVLTL 473  
Qy 556 YHVPVSGCMD-CSESDHRI-FMNTCNPSLSLQOQLFHTNSTVLEKF 600  
Db 474 QHVNNOCLDKATBEDSQVPSIRDCS-GSRSQOMLLR--NVTLPBIF 517

## RESULT 10

PCT-US94-02552-19  
Sequence 19, Application PC/TUS9402552  
GENERAL INFORMATION:  
APPLICANT: Elhammer, Ake P.  
APPLICANT: Homa, Fred L.  
TITLE OF INVENTION: A cloned DNA Encoding a UDP-galNAC:  
TITLE OF INVENTION: Polypeptide, N-Acetylglactosaminyltransferase  
NUMBER OF SEQUENCES: 19  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: The Upjohn Company, Corp. Intellectual  
ADDRESSEE: Property Law  
STREET: 301 Henrietta Street  
CITY: Kalamazoo  
STATE: Michigan  
COUNTRY: USA  
ZIP: 49001  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US94/02552  
FILING DATE:  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Darnley Jr., James D.  
REGISTRATION NUMBER: 33,673  
REFERENCE/DOCKET NUMBER: 4755.P CP  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 616-385-5210  
TELEFAX: 616-385-6897  
TELEX: 224401  
INFORMATION FOR SEQ ID NO: 19:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 517 amino acids  
TYPE: amino acid

STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
PCT-US94-02552-19

Query Match 34.0%; Score 115; DB 5; Length 517;  
Best Local Similarity 43.6%; Pred. No. 8.3e-111;  
Matches 230; Conservative 89; Mismatches 170; Indels 38; Gaps 16;

QY 88 GNGEQRP--YPTDAERVDQAYRENGFNIVYSDKISLNSLPDIRHPCNSKRYLETLP 145  
DB 15 GPGMGKPVVLPKEDQKMEKMFKNQFNLMASEMIALNSLPDVRLEGCKTKVYPNLP 74  
QY 146 NTSIIIFHNEGSSILRTVSHVSNRSPPELVAVIVDDPSREHLKPELIDYM-ALFP 204  
DB 75 TTSVIVFHNEMASTLRTVSHVSNRSPPELVAVIVDDPSREHLKPELIDYM-ALFP 134  
QY 205 SVRIATKKEGLIRTRMLGASVATGDTFTLDSHCENYVNLPLDIRIARNTIYCP 264  
DB 135 PVHIVMEQSGILIRALKGAASVKGQVITFLDAHCECTGWLPELARIKHDRKTYVCP 194  
QY 265 MIDVIDHDDFRYETQAG-DAMRGAFDMEWYKRIPIPP--ELQKADSPSPSSPVNAG 320  
DB 195 IIDVISDITFEY--MAGSDMTYGGFNKMLNFRWYVPQREMDRKGRTLPVPTPTMAGG 252  
QY 321 LFAVDRKFMELGSDYDGLIHWGEOYEISFKVWMCGRMEDIPCSRVGHYRKTYVYK 380  
DB 253 LFSIDROYFOEIGTYDAGMDIWGENLEISFRWOCGTEIYVCSHVGHVFRATPYTF 312  
QY 381 PAGVS--LADNLKRAVAVMDVAYEYIYQRRPEYRHLASDVAVOKKLRSLSNCKSPKWF 438  
DB 313 PGGTGQIINKNRRLAEVMDDEFKNFYIISPGYTKVDYDISRLGRHLQCRPSPWY 372  
QY 439 MTKIAMD--LPKRYPPVEPPAAAMGEIRNVGTGLCADTKHAGLSPLRLGCVRGEBAA 496  
DB 373 LENIYPSQIPRYH-----FSLGEIRNVETNOCLDNMARKENKVGIFNC-HGMG--- 421  
QY 497 WNNQVFTFMWRDIRPDQHTKKFCFDAISHTSPTLYDCHSMKGNQMLKRYKDK-TL 555  
DB 422 --GNQVFSYANKKIRDD-----LCLDYSKLNGLPYTMKCHHKGQWMBEYDPVLT 473  
QY 556 YHVPVSGCMD-CSESDHRI-FMNTCNPSLSLTQOMLFETHNSTVLEKF 600  
DB 474 QHVNNSQCLDKATDEDSQVPSIRDGS-GSRQOWILR--NVTLPETIF 517

RESULT 11  
US-08-967-508-9  
Sequence 9, Application US/08967508  
Patent No. 5910570  
GENERAL INFORMATION:  
APPLICANT: The Upjohn Company  
APPLICANT: FOR U.S. PURPOSES ONLY: Elihammer, Ake P. and Homa, Fred L.  
TITLE OF INVENTION: A Cloned DNA Encoding a UDP-GALNAC:  
TITLE OF INVENTION: Polypeptide, N-Acetylglactosaminyltransferase  
NUMBER OF SEQUENCES: 19  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Pharmacia & Upjohn Company, Intellectual  
STREET: 301 Henrietta Street  
CITY: Kalamazoo  
STATE: Michigan  
COUNTRY: USA  
ZIP: 49001  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/967,508  
FILING DATE:  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Pharmacia & Upjohn Company, Intellectual  
CLASSIFICATION: 435

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/602,830  
FILING DATE: 13 No. 5910570member 1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Darnley Jr., James D.  
REGISTRATION NUMBER: 33,673  
REFERENCE/DOCKET NUMBER: 4755.P CP  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 616-833-2210  
TELEFAX: 616-833-8897  
TELEX: 224401

INFORMATION FOR SEQ ID NO: 9:

SEQUENCE CHARACTERISTICS:  
LENGTH: 559 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-967-508-9

Query Match 34.0%; Score 115; DB 2; Length 559;  
Best Local Similarity 43.6%; Pred. No. 9.5e-111;  
Matches 230; Conservative 89; Mismatches 170; Indels 38; Gaps 16;

QY 88 GNGEQRP--YPTDAERVDQAYRENGFNIVYSDKISLNSLPDIRHPCNSKRYLETLP 145  
DB 57 GPGMGKPVVLPKEDQKMEKMFKNQFNLMASEMIALNSLPDVRLEGCKTKVYPNLP 116  
QY 146 NTSIIIFHNEGSSILRTVSHVSNRSPPELVAVIVDDPSREHLKPELIDYM-ALFP 204  
DB 117 TTSVIVFHNEMASTLRTVSHVSNRSPPELVAVIVDDPSREHLKPELIDYM-ALFP 176  
QY 205 SVRIATKKEGLIRTRMLGASVATGDTFTLDSHCENYVNLPLDIRIARNTIYCP 264  
DB 177 PVHIVMEQSGILIRALKGAASVKGQVITFLDAHCECTGWLPELARIKHDRKTYVCP 236  
QY 265 MIDVIDHDDFRYETQAG-DAMRGAFDMEWYKRIPIPP--ELQKADSPSPSSPVNAG 320  
DB 237 IIDVISDITFEY--MAGSDMTYGGFNKMLNFRWYVPQREMDRKGRTLPVPTPTMAGG 294  
QY 497 WNNQVFTFMWRDIRPDQHTKKFCFDAISHTSPTLYDCHSMKGNQMLKRYKDK-TL 555  
DB 295 LFSIDROYFOEIGTYDAGMDIWGENLEISFRWOCGTEIYVCSHVGHVFRATPYTF 354  
QY 381 PAGVS--LADNLKRAVAVMDVAYEYIYQRRPEYRHLASDVAVOKKLRSLSNCKSPKWF 438  
DB 355 PGGTGQIINKNRRLAEVMDDEFKNFYIISPGYTKVDYDISRLGRHLQCRPSPWY 414  
QY 439 MTKIAMD--LPKRYPPVEPPAAAMGEIRNVGTGLCADTKHAGLSPLRLGCVRGEBAA 496  
DB 415 LENIYPSQIPRYH-----FSLGEIRNVETNOCLDNMARKENKVGIFNC-HGMG--- 463  
QY 497 WNNQVFTFMWRDIRPDQHTKKFCFDAISHTSPTLYDCHSMKGNQMLKRYKDK-TL 555  
DB 464 --GNQVFSYANKKIRDD-----LCLDYSKLNGLPYTMKCHHKGQWMBEYDPVLT 515  
QY 556 YHVPVSGCMD-CSESDHRI-FMNTCNPSLSLTQOMLFETHNSTVLEKF 600  
DB 516 QHVNNSQCLDKATDEDSQVPSIRDGS-GSRQOWILR--NVTLPETIF 559

RESULT 12  
US-08-967-506-9  
Sequence 9, Application US/08967506  
Patent No. 6096512  
GENERAL INFORMATION:  
APPLICANT: The Upjohn Company  
APPLICANT: FOR U.S. PURPOSES ONLY: Elihammer, Ake P. and Homa, Fred L.  
TITLE OF INVENTION: A Cloned DNA Encoding a UDP-GALNAC:  
TITLE OF INVENTION: Polypeptide, N-Acetylglactosaminyltransferase  
NUMBER OF SEQUENCES: 19  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Pharmacia & Upjohn Company, Intellectual



QY 439 MTKIAMD--LPKXPVYBPAPAAWGEIRNVGTGACDTKHGALGSLRLBECVARGBEAA 496  
DB 415 LENIYDPSQJPRHY-----FSLGEIRNVETNOCLDMARKENKEIGENC-HGWG-- 463  
QY 497 WNNWQVFTFWREDIRBGPQHTKKFCFDAISHTSPVLYDCHSGMKNQMLKRYAKDK-TL 555  
DB 464 --GNQVFSYTNKRIKIRDD-----LCIDVSKLNGPVTMLKCHLKNQMLMEYDPVLTLL 515  
QY 556 YHPVSGSCMD-CSESDEHRI-FMNTCNBSLTOQWLFHTNSTVLKFP 600  
DB 516 QHVNNOCLDKATDEDSQVPSIRDOS-GSRGQWMLLR--NVTLEIF 559

RESULT 14  
US-09-217-306B-2  
Sequence 2, Application US/09217306B  
Patent No. 6465220  
GENERAL INFORMATION:  
APPLICANT: Haasan, Helle  
APPLICANT: Clausen, Henrik  
TITLE OF INVENTION: Glycosylation Using GalNac-T4 Transferase  
FILE REFERENCE: 8850\*1  
CURRENT APPLICATION NUMBER: US/09/217,306B  
CURRENT FILING DATE: 1998-12-21  
NUMBER OF SEQ ID NOS: 25  
SOFTWARE: Patentin version 3.1  
SEQ ID NO 2  
LENGTH: 578  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-217-306B-2

Query Match 31.3%; Score 1026; DB 4; Length 578;  
Best Local Similarity 39.4%; Pred. No. 3,9e-101;  
Matches 221; Conservative 98; Mismatches 204; Indels 38; Gaps 16;

QY 50 AAGGSSHRQKTFPLDGOGLKMDHKEAIR--DAQVNGEQQGPRYMTDAE--RV 104  
DB 35 SAGAG-RARELGSRRSLDQKNTEDLSRPLKPPADSRALGEWQKASKQLNDELKQ 93  
QY 105 DQARENGFNIVYDSTSLNSLSPDIRHPCNSKRY-LETPNTSIIIFHNBSGMSILR 163  
DB 94 BELIERAINIYLSDRISLHRIEDKRYECKSQFNYRTLPSSVIAFNEAMSTLR 153  
QY 164 TVHSVLNRPEPELVAEIVLVDFSDREHLKPLEDYMLFPSVRIILRTKKEGILRTML 223  
DB 154 TIHSVLSPAVLKEIILVDDLSDRVYLKQLETYSINDRVALLIRNKREGILVRAALI 213  
QY 224 GASVATGDVITFLDSHCANVNLPLIDRIARNKTIIVCPMIDVIDHDFRYETQAGDA 283  
DB 214 GATPATGDVITFLYCHCEGNSGMLPELIERIGRYETAIVACVIDIINTFEFYMQGEP 273  
QY 284 MRGAFDEMYKRIPIPEL--QKADPSDPRESVMAAGLFAVDKRWMLGVDPLGI 341  
DB 274 MGGFDWRLTFQWHSVPRQERDRIRISIDPIRSPTMAGGLFAVSKTFQYLGTDGMEV 333  
QY 342 WGSQVEYISFKVMWCGGEMEDIPCSRVGHIYKRYVYVPAVGLSARNLKVAEVMDEX 401  
DB 334 WGSNLSLSPVWOCGKLEIHPGSHGVHVPKAPARP--NFLQNTAAALVWMDY 390  
QY 402 AEYIYQRPPEYRHLASGDVAQKLRSLNSKSPKFWMTKIAMDLPKRYVPEBPAAW- 460  
DB 391 KEHFNYNRPARKAYGDISERKLRERLCKSPDWYLNKVN--FPNLAHVEDBRP--GMH 445  
QY 461 GEIRNVG-TGLCAD--TKHGALGSLRLBECVARGBEAANNNQVFTFWREDIRBGP 516  
DB 446 GAISRGISSECLDYNBPNNPTGANLSLFGC-HGQG--GNQFEYTSNKEIRFNS- 498  
QY 517 QHTKKFCFDAISHTSPVLYDCHSGMKNQMLKRYAKDK-TL 555  
DB 499 --VTLECAVEPEQKNYVGMQNCPRDGFVVPANIIWHFREDGTIFHPHSGCLSLAVRTP 556

QY 570 DHRIFMNTCNBSLSLTQWLF 590  
DB 557 RPDVQKRTCDALDKNQIWSFE 577

RESULT 15  
US-09-217-306B-10  
Sequence 10, Application US/09217306B  
Patent No. 6465220  
GENERAL INFORMATION:  
APPLICANT: Haasan, Helle  
APPLICANT: Clausen, Henrik  
TITLE OF INVENTION: Glycosylation Using GalNac-T4 Transferase  
FILE REFERENCE: 8850\*1  
CURRENT APPLICATION NUMBER: US/09/217,306B  
CURRENT FILING DATE: 1998-12-21  
NUMBER OF SEQ ID NOS: 25  
SOFTWARE: Patentin version 3.1  
SEQ ID NO 10  
LENGTH: 560  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Interferon fragment-galNac-T4 predicted coding region  
NAME/KEY: PEPTIDE  
LOCATION: (1)..(25)  
OTHER INFORMATION: human interferon signal peptide  
FEATURE:  
NAME/KEY: MISC\_FEATURE  
LOCATION: (37)..(560)  
OTHER INFORMATION: galNac-T4  
US-09-217-306B-10

Query Match 31.2%; Score 1024; DB 4; Length 560;  
Best Local Similarity 41.0%; Pred. No. 6,1e-101;  
Matches 213; Conservative 88; Mismatches 183; Indels 36; Gaps 14;

QY 90 GEOGRPYPM---TDAERVDOARENGFNIVYDSTSLNSLSPDIRHPCNSKRY-LETL 144  
DB 57 GEMWQKASKQLNDELKQBELIERAINIYLSDRISLHRIEDKRYECKSQFNYRTLP 116  
QY 145 PNTSIIIFHNBSGMSILRTVHSVLNRSPEPELVAEIVLVDFSDREHLKPLEDYMLFP 204  
DB 117 PPTSIVIAFNEAMSTLRIHISVLTSPAVLKEIILVDDLSDRVYLKQLETYSINLD 176  
QY 205 SVRIIRTKKREGILRTMLGASVATGDVITFLDSHCANVNLPLIDRIARNKTIIVCP 264  
DB 177 RVRLIRTKKREGILRAVLIGATPATGDVITFLYCHCEGNSGMLPELIERIGRYETAIVCP 236  
QY 265 MIDVIDHDDRYETQAGDAMRGAFDEMYKRIPIPEL--QKADPSDPRESVMAAGLF 322  
DB 237 VIDIDINTFEFYMQGEPMDRLTFQWHSVPRQERDRIRISIDPIRSPTMAGGLF 296  
QY 323 AVDRKFWMLGSDYDPLGIEINGEYETISFKVMWCGGEMEDIPCSRVGHIYKRYVYVPA 382  
DB 297 AVSKRYFQYLGTDGMEVWCGENLISLSPVWOCGKLEIHPGSHGVHVPKAPARP- 355  
QY 383 GVSIAIRLKVAVRWMDEYAEYIYQRPPEYRHLASGDVAQKLRSLNSKSPKFWMTKI 442  
DB 356 --NFLQNTAAALVWMDYEHFNRPARKAYGDISERKLRERLCKSPDWYLNKVN- 413  
QY 443 AMDLPKRYVPEBPAAW-GEIRNVG-TGLCAD--TKHGALGSLRLBECVARGBEAAW 497  
DB 414 --FPNLAHVEDBRP--GMHGAISRGISSECLDYNBPNNPTGANLSLFGC-HGQG-- 463  
QY 498 NNNQVFTFWREDIRBGPQHTKKFCFDAISHTSPVLYDCHSGMKNQMLKRYAKDK 553  
DB 464 --GNQFEYTSNKEIRFNS--VTLECAVEPEQKNYVGMQNCPRDGFVVPANIIWHFREDG 519  
QY 554 TVHNPVSGSCMD--SESDHRIFMNTCNBSLSLTQWLF 590

Db 520 TIFPHSGTCLSAVRTPEGRPDVQRTCDALDKXQIMFE 559

## RESULT 16

US-09-795-926-33

Sequence 33, Application US/09795926

Patent No. 6555669

GENERAL INFORMATION:

APPLICANT: Donoho, Gregory

APPLICANT: Hilbun, Erin

APPLICANT: Turner, C. Alexander Jr.

APPLICANT: Friedlich, Glenn

APPLICANT: Abuin, Alejandro

APPLICANT: Zambrowicz, Brian

APPLICANT: Sands, Arthur T.

APPLICANT: Walke, D. Wade

APPLICANT: Wilganowski, Nathaniel L.

APPLICANT: Hu, Yi

APPLICANT: Kieke, James Alvin

APPLICANT: Potter, David George

TITLE OF INVENTION: NOVEL HUMAN TRANSFERASE PROTEINS AND

TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING THE SAME

FILE REFERENCE: LEX-0144-USA

CURRENT APPLICATION NUMBER: US/09/795,926

PRIOR APPLICATION NUMBER: US 60/185,920

PRIOR FILING DATE: 2000-02-29

PRIOR APPLICATION NUMBER: US 60/186,558

PRIOR FILING DATE: 2000-03-02

PRIOR APPLICATION NUMBER: US 60/191,849

PRIOR FILING DATE: 2000-03-24

NUMBER OF SEQ ID NOS: 47

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 33

LENGTH: 240

TYPE: PRT

ORGANISM: homo sapiens

US-09-795-926-33

Query Match 30.2%; Score 990; DB 4; Length 240;

Best Local Similarity 97.9%; Pred. No. 6.8e-98;

Matches 189; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 MRREKRLQAVLALVLAIVLPVNGMALYREROPDTPGSSGAAYAPAAQSGSHSRQK 60

Db 1 MRREKRLQAVLALVLAIVLPVNGMALYREROPDTPGSSGAAYAPAAQSGSHSRQK 60

QY 61 KTFPLGDOGQKLDKMDHDEKAIIRDAQVNGEGRPYPTMDAERVDQAYRENGFNIIYSDK 120

Db 61 KTFPLGDOGQKLDKMDHDEKAIIRDAQVNGEGRPYPTMDAERVDQAYRENGFNIIYSDK 120

QY 121 ISLRSLEPDIRHPNCSKRYETLPTNTSIIIPFNHSGSSLLRTVHSLVNSPPELVAEI 180

Db 121 ISLRSLEPDIRHPNCSKRYETLPTNTSIIIPFNHSGSSLLRTVHSLVNSPPELVAEI 180

QY 181 VLVDPSDRHLK 193

Db 181 VLVDPSDRGRIR 193

RESULT 17

US-09-795-926-37

Sequence 37, Application US/09795926

Patent No. 6555669

GENERAL INFORMATION:

APPLICANT: Donoho, Gregory

APPLICANT: Hilbun, Erin

APPLICANT: Turner, C. Alexander Jr.

APPLICANT: Friedlich, Glenn

APPLICANT: Abuin, Alejandro

APPLICANT: Zambrowicz, Brian

APPLICANT: Sands, Arthur T.

APPLICANT: Walke, D. Wade

APPLICANT: Wilganowski, Nathaniel L.

APPLICANT: Hu, Yi

APPLICANT: Kieke, James Alvin

APPLICANT: Potter, David George

TITLE OF INVENTION: NOVEL HUMAN TRANSFERASE PROTEINS AND

TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING THE SAME

FILE REFERENCE: LEX-0144-USA

CURRENT APPLICATION NUMBER: US/09/795,926

PRIOR APPLICATION NUMBER: US 60/185,920

PRIOR FILING DATE: 2000-02-29

PRIOR APPLICATION NUMBER: US 60/186,558

PRIOR FILING DATE: 2000-03-02

PRIOR APPLICATION NUMBER: US 60/191,849

PRIOR FILING DATE: 2000-03-24

APPLICANT: Walke, D. Wade

APPLICANT: Wilganowski, Nathaniel L.

APPLICANT: Hu, Yi

APPLICANT: Kieke, James Alvin

TITLE OF INVENTION: NOVEL HUMAN TRANSFERASE PROTEINS AND

TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING THE SAME

FILE REFERENCE: LEX-0144-USA

CURRENT APPLICATION NUMBER: US/09/795,926

PRIOR APPLICATION NUMBER: US 60/185,920

PRIOR FILING DATE: 2000-02-29

PRIOR APPLICATION NUMBER: US 60/186,558

PRIOR FILING DATE: 2000-03-02

PRIOR APPLICATION NUMBER: US 60/191,849

PRIOR FILING DATE: 2000-03-24

NUMBER OF SEQ ID NOS: 47

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 37

LENGTH: 209

TYPE: PRT

ORGANISM: homo sapiens

US-09-795-926-37

Query Match 30.1%; Score 988; DB 4; Length 209;

Best Local Similarity 100.0%; Pred. No. 8.8e-98;

Matches 189; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MRREKRLQAVLALVLAIVLPVNGMALYREROPDTPGSSGAAYAPAAQSGSHSRQK 60

Db 1 MRREKRLQAVLALVLAIVLPVNGMALYREROPDTPGSSGAAYAPAAQSGSHSRQK 60

QY 61 KTFPLGDOGQKLDKMDHDEKAIIRDAQVNGEGRPYPTMDAERVDQAYRENGFNIIYSDK 120

Db 61 KTFPLGDOGQKLDKMDHDEKAIIRDAQVNGEGRPYPTMDAERVDQAYRENGFNIIYSDK 120

QY 121 ISLRSLEPDIRHPNCSKRYETLPTNTSIIIPFNHSGSSLLRTVHSLVNSPPELVAEI 180

Db 121 ISLRSLEPDIRHPNCSKRYETLPTNTSIIIPFNHSGSSLLRTVHSLVNSPPELVAEI 180

QY 181 VLVDPSDR 189

Db 181 VLVDPSDR 189

RESULT 18

US-10-140-002-196

Sequence 196, Application US/10140002

Patent No. 6725730

GENERAL INFORMATION:

APPLICANT: Baker, Kevin P.

APPLICANT: Beresini, Maureen

APPLICANT: Deforge, Laura

APPLICANT: Desnoyers, Luc

APPLICANT: Filvarioff, Ellen

APPLICANT: Gao, Wei-Qiang

APPLICANT: Gerritsen, Mary B.

APPLICANT: Goddard, Audrey

APPLICANT: Godowski, Paul J.

APPLICANT: Gurney, Austin L.

APPLICANT: Sherwood, Steven

APPLICANT: Smith, Victoria

APPLICANT: Stewart, Timothy A.

APPLICANT: Tamas, Daniel

APPLICANT: Watanabe, Colin K

APPLICANT: Wood, William

APPLICANT: Zhang, Zemin

TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC

TITLE OF INVENTION: ACIDS ENCODING THE SAME

FILE REFERENCE: P3330R1C59

CURRENT APPLICATION NUMBER: US/10/140,002

CURRENT FILING DATE: 2002-05-06

Prior Application removed - See Palm or File Wrapper





```

OY      202  FSPSRLIRTKKREGIIRFRMGASVATGDTVTLFSLHCBAANVMYPLRLDIRAARXTI 261
Db      242  QFSYKIVRQGRKRLITRLLGATATATLETTLFDHABCECFGLMBPLIARIAENYAV 301
OY      262  VCPMDIVDHDHDFRYETQA--GDAMRGAFDWEMY--KRIP1PEIQAQPSDPESPV 316
Db      302  VSPDIASIDLNTFEFNKPSRPGYSNNHRGNFDMWLSLPGWESLPHDEKQRDRDEYPIKTP 361
OY      317  MAGGIFAADRKKFWELGCTDPCGLETWGGGQYELSEPKVMCGGMEDIPCSRVGHITRYV 376
Db      362  FAGGFFSISKEYFEYIGSYDEEMELWGGNIEMSPFWOCGGGLELMPCSVGHVFRSKS 421
OY      377  FYKVNAGYS-LARNLKRVAVNMWMDAEAYETIYORPE----YRHLISADVAVQKGLSSLN 431
Db      422  PHSFPGKQIVIRKNQVRRLAEVMMDEYKLFYRRNTDAAKIVKXAKAGDLSKREIRRLR 481
OY      432  CKSPKCFMTKIAMLDPKFYRPVEPAAAMGEIRNVTGTCADT-KIGALGSLPLRLGCVR 490
Db      482  CKNFMYMLNTI---YPEYVVPDLNPVTS-GYIKSVGQPLCLDVGENNGGKPLIMTTC-H 536
OY      491  GRGEAAMNNMQUFTTWREDDIRPDPQHT--KKCFDAI-----SHTSPVTLY 536
Db      537  GLG-----GNQYFEXSAQHEIR-----HNIOKELCHAAQGLVOLKACTYKGGKTVVT-- 584
OY      537  DCHSMKGNQIATCYRDKTLYHPVSGSCMPCSSDHRIFMNTCPSSSLTOQMLEHNTN 593
Db      585  -----GEOIMEIQXQDULYNPLKMCIS-ANGEHPSLV-SCNPSDPLQAMWISQND 633

```

```

Oy      98 MTDARVVOAAYENGFNIVSDKISLNRSLPDIRHNCNSKRLETLPNTSIIIPHNNEG 157
Ob      1 MTDARVVOAARENGFNIVSDKISLNRLPIDIRHNCNSKRLETLPNTSIIIPHNNEG 60

; RESULT 23
US-09-795-926-23
; Sequence 23, Application US/09795926
Patent No. 6555669
; GENERAL INFORMATION:
APPLICANT: Donoho, Gregory
APPLICANT: Hilpun, Erin
APPLICANT: Turner, C. Alexander Jr.
APPLICANT: Friedrich, Glenn
APPLICANT: Abuin, Alejandro
APPLICANT: Zambrowski, Brian
APPLICANT: Sands, Arthur T.
APPLICANT: Walke, D. Wade
APPLICANT: Wilganowski, Nathaniel L.
APPLICANT: Hu, Yi
APPLICANT: Kieker, James Alvin
APPLICANT: Potter, David George
TITLE OF INVENTION: NOVEL HUMAN TRANSFERASE PROTEINS AND
TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING THE SAME
FILE REFERENCE: LEX-0144-USA
CURRENT APPLICATION NUMBER: US/09/795,926
CURRENT FILING DATE: 2001-02-28
PRIOR APPLICATION NUMBER: US 60/185,920
PRIOR FILING DATE: 2000-02-29
PRIOR APPLICATION NUMBER: US 60/186,558
PRIOR FILING DATE: 2000-03-02
PRIOR APPLICATION NUMBER: US 60/191,849
PRIOR FILING DATE: 2000-03-24
NUMBER OF SEQ ID NOS: 47
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 23
LENGTH: 224
TYPE: PRT
ORGANISM: homo sapiens
US-09-795-926-23

Query Match          24.5%; Score 804; DB 4; Length 224;
Best Local Similarity 98.7%; Pred. No. 6, 7e-78;
Matches 154; Conservative 1; Mismatches 1; Indels 0; Gaps 0
```

Qy 158 WSLSLATVTSYLNRSPELVVAETVLVDVDFSRHKKPLBQYMALPSVRLIRTKKEGL 211  
Db 61 WSSLATVTSYLNRSPELVVAETVLVDVDFSRHKKPLBQYMALPSVRLIRTKKEGL 120

```

RESULT 24
US-09-973-457-2
: Sequence 2, Application US/09973457
: Patent No. 6703230
: GENERAL INFORMATION:
: APPLICANT: Kapeller-Libermann, Roana
: TITLE OF INVENTION: 47174, A NOVEL HUMAN GLYCOSYLTRANSFERASE
: FILE REFERENCE: 10448-099001
: CURRENT APPLICATION NUMBER: US/09/973,457
: CURRENT FILING DATE: 2001-10-09
: PRIOR APPLICATION NUMBER: 60/238,849
: PRIOR FILING DATE: 2000-10-06
: NUMBER OF SEQ ID NOS: 6
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 2
: LENGTH: 603
: TYPE: PRF
: ORGANISM: Homo sapiens
US-09-973-457-2

```

Query Match	22.7%	Score 744;	DB 4;	Length 603;
Best Local Similarity	32.3%	Pred. No. 1e-70;		
Matches	202;	Conservative	86;	Mismatches 220; Indels 108; Gaps 22;
Qy	3	RKEKRLIQAAVALVALVLLPNVGLMALYERBDPTGGSGAAVADAG-QGSHSROKK	61	
Db	5	RKLTLLTNVILVPGVIG-----LFSVVCRLQ-----GRSOELVRRIVSGDRRVRSHAK	53	
Qy	62	TFPLFGDQOKLDM--HKEAIIIRDAQGV-----GNGEGGRPYM-TDAERDQA	107	
Db	54	VGTIGDDEBALIQRDLHEVYVYNQNLGAPIGLVEGPGIGGGGLAATLRDDQBAEGK	113	
Qy	108	YRENGFNIIYVSDKISLNRSLPDRIHPNCSKRYLETLPNTSIIIPFNEGSSLLRTVHS	167	
Db	114	YEEGYNAQJSDRISLDRSIPDYPRKRCQMSYADLPQVSVVFIFPNEGALSTVLRSHS	173	
Qy	168	VLRNSPEELVAEIVLYNDPFDREHLKKPLBDYM-ALPSS-VRLTRTKRREGI.PTRMLGA	225	
Db	174	VVNTTPEQLKEVLLVYDNSDNEELKKNLQOYNAKRPGLAKIVRNSRRBSLIPARIQGW	233	
Qy	226	SVATGADVITFEDSHCEANVMPLPELDRIARNRKTIYCPMIDVIDHDHDFRETQGDAMR	285	
Db	234	KAAPAIVGVGFDAVEFNTGMABALSRINEDRRRIYLPAINIKYSTFEVQ-QYANAAH	292	
Qy	286	GAFPMENYTKRIPIPEP--LQKADPSPDPFSSPYWAGGLFAVDRKMFELGSDYDGLGIWG	343	
Db	293	G-YVMGIMCMYI-LPDDMLDRGESAIPITPAIMIGSFVVDREYFDIGILPGEIMEYVG	350	
Qy	344	GEOYEISFKYMMCGGRREDIPCSRGVGIYKRYVYKVPAGVSLARNIKRAVAEVMDEYAE	403	
Db	351	GENVELGMARWQCGSGSEVLPCSRVAIHETRTKRYNNDIIVYAKRNMLRAAEVWMDPFKS	410	
Qy	404	YTIYOR--RPEYRHLASGDAVAOQKRRSSLNCSFKKFMFKLIAMDEKFTYPP--VEBPAA	458	
Db	411	HYVYAAMNIIPMSNPVDVGDGVSERLATRQRIKCRSFKYLLENV-----YPERKVVYNTL	463	
Qy	459	AMGEIRN-VENGLCADPFRKHALGSPLEBECVGRGGAANNMQVFTFRREDIRPDGPQ	517	
Db	464	TYGVRVRSKSKSAYCLD--QGA-----ED---GD---486		
Qy	518	HTKPCFCDALSHTSPTVLLDYCHSNKGNQMLKRRD-----KTLVHPVSGSCMDCSE	568	

Db 487 -----RAILYPCHGM-SGQLVRYSDAGLLQLGLGSTAFLPDS-KCLVDDG 530  
QY 569 SDHRIFMATCNPSLSLTOQWLFEHTNS 594  
Db 531 TGRMPTLKKECDVARPTORLMDFTOS 556

RESULT 25

US-09-270-767-45334  
; Sequence 45334, Application US/09270767  
; Patent No. 6703491  
; GENERAL INFORMATION:  
; APPLICANT: Homburger et al.  
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster  
; FILE REFERENCE: File Reference: 7326-094  
; CURRENT APPLICATION NUMBER: US/09/270,767  
; CURRENT FILING DATE: 1999-03-17  
; NUMBER OF SEQ ID NOS: 62517  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 45334  
; LENGTH: 289  
; TYPE: PRT  
; ORGANISM: Drosophila melanogaster  
; FEATURE:  
; OTHER INFORMATION: Xaa means any amino acid  
US-09-270-767-45334

Query Match 20.6%; Score 675; DB 4; Length 289;

Best Local Similarity 50.2%; Pred. No. 8.2e-64; Matches 131; Conservative 45; Mismatches 77; Indels 8; Gaps 4;

QY 111 NGFNIVYSDKISLNRSLEPDIRHPNCNSKRYLETLPNTSIIIPFNHGWSSSLRTVHSYLN 170  
Db 34 NRFNGEASDALPSNRDIDPTNPMCRCTKKYREDLPETSVITTFHNEARSTLLRTIVSYLN 93  
QY 171 RSPPELVAEIVLVDDPSDREHLKKPLEDYMALFESVRIILRTKREGLIRTRMLGASVATG 230  
Db 94 RSPPEHLIREIVLVDDYSD--HPEDGLE--LAKIDKRVIRNDKREGLVRSRVKGDAAVS 149  
QY 231 DVITFLDSHCANVNWLPPLLDRIARNRKTIVCPMIDVIDHDDPFRYETQAGDAMRGAPDW 290  
Db 150 SVLTFLDSHVCENMWELEPLERVAEDPTRVVCPVIDVISMDNFOYIGASAD-LRGGPDW 208  
QY 291 EMYKRIPIPE--LQKADPSDPESPVMAGLFAVDKRWFWELGGYDPGLBIWGEQY 347  
Db 209 NLIFKWEYLSPESEAMRNDDPTTAIRTEMINGGLFVIDKAYFNKLKGYDMKMDVWGENTL 268  
QY 348 EISFRVMCGGRMEDIPCSR 368  
Db 269 EISFRVMCGGSLRIIPCSR 289

Search completed: November 22, 2004, 13:44:43  
Job time : 23 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 22, 2004, 13:37:52 ; Search time 57 Seconds

(Without alignments)  
3746.297 Million cell updates/sec

Title: US-10-001-851-2

Perfect score: 3278

Sequence: 1 MRKERRLLQAVLVLAALV.....TQGMLEPHNTSYLKEKRN 603

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1570615 seqs, 354127592 residues

Total number of hits satisfying chosen parameters: 1570615

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database : Published Applications AA:\*

1: /cgn2\_6/ptodata/1/pubppaa/US07\_PUBCOMB.pep:\*

2: /cgn2\_6/ptodata/1/pubppaa/PCR\_NEW\_PUB.pep:\*

3: /cgn2\_6/ptodata/1/pubppaa/US06\_NEW\_PUB.pep:\*

4: /cgn2\_6/ptodata/1/pubppaa/US06\_PUBCOMB.pep:\*

5: /cgn2\_6/ptodata/1/pubppaa/US07\_NEW\_PUB.pep:\*

6: /cgn2\_6/ptodata/1/pubppaa/PCRUS\_PUBCOMB.pep:\*

7: /cgn2\_6/ptodata/1/pubppaa/US08\_NEW\_PUB.pep:\*

8: /cgn2\_6/ptodata/1/pubppaa/US08\_PUBCOMB.pep:\*

9: /cgn2\_6/ptodata/1/pubppaa/US09A\_PUBCOMB.pep:\*

10: /cgn2\_6/ptodata/1/pubppaa/US09B\_PUBCOMB.pep:\*

11: /cgn2\_6/ptodata/1/pubppaa/US09C\_PUBCOMB.pep:\*

12: /cgn2\_6/ptodata/1/pubppaa/US09C\_NEW\_PUB.pep:\*

13: /cgn2\_6/ptodata/1/pubppaa/US10A\_PUBCOMB.pep:\*

14: /cgn2\_6/ptodata/1/pubppaa/US10B\_PUBCOMB.pep:\*

15: /cgn2\_6/ptodata/1/pubppaa/US10C\_PUBCOMB.pep:\*

16: /cgn2\_6/ptodata/1/pubppaa/US10D\_PUBCOMB.pep:\*

17: /cgn2\_6/ptodata/1/pubppaa/US10\_NEW\_PUB.pep:\*

18: /cgn2\_6/ptodata/1/pubppaa/US11\_NEW\_PUB.pep:\*

19: /cgn2\_6/ptodata/1/pubppaa/US60\_NEW\_PUB.pep:\*

20: /cgn2\_6/ptodata/1/pubppaa/US60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3278	100.0	603	9	US-09-795-926-43
2	3278	100.0	603	13	US-10-001-851-2
3	3278	100.0	603	14	US-10-364-774-43
4	3242.5	98.9	631	9	US-09-795-926-41
5	3242.5	98.9	631	14	US-10-364-774-41
6	2771	84.5	506	9	US-09-795-926-31
7	2771	84.5	506	14	US-10-364-774-31
8	2746.5	83.8	535	9	US-09-795-926-29
9	2746.5	83.8	535	14	US-10-364-774-29
10	1877	57.3	366	9	US-09-795-926-39
11	1877	57.3	366	14	US-10-364-774-39
12	1392	42.5	276	14	US-10-292-896-3
13	1370	41.8	269	9	US-09-795-926-27

14	269	14	US-10-364-774-27	Sequence 27, App1
15	321	9	US-09-795-926-35	Sequence 35, App1
16	321	14	US-10-364-774-35	Sequence 35, App1
17	561	9	US-09-925-301-1006	Sequence 1006, App
18	559	13	US-10-001-851-24	Sequence 24, App1
19	559	13	US-10-001-851-21	Sequence 21, App1
20	559	14	US-10-205-219-36	Sequence 36, App1
21	559	14	US-10-205-219-76	Sequence 76, App1
22	559	13	US-10-001-851-22	Sequence 22, App1
23	559	13	US-10-001-851-23	Sequence 23, App1
24	556	14	US-10-292-896-62	Sequence 20, App1
25	626	13	US-10-001-851-27	Sequence 62, App1
26	581	15	US-10-074-527-2	Sequence 27, App1
27	578	14	US-10-292-896-118	Sequence 118, App
28	578	13	US-10-074-527-8	Sequence 8, App1
29	657	14	US-10-341-434-32	Sequence 32, App1
30	558	14	US-10-292-896-66	Sequence 66, App1
31	558	15	US-10-433-256-6	Sequence 6, App1
32	581	13	US-10-074-527-2	Sequence 2, App1
33	581	15	US-10-085-198-122	Sequence 122, App
34	581	15	US-10-433-256-4	Sequence 4, App1
35	581	16	US-10-789-241-44	Sequence 44, App1
36	240	9	US-09-795-926-33	Sequence 33, App1
37	240	14	US-10-364-774-33	Sequence 33, App1
38	612	13	US-10-001-851-25	Sequence 25, App1
39	209	9	US-09-795-926-37	Sequence 37, App1
40	209	14	US-10-364-774-37	Sequence 37, App1
41	581	14	US-10-292-896-58	Sequence 58, App1
42	552	9	US-09-815-028-2	Sequence 2, App1
43	552	14	US-10-028-072-196	Sequence 196, App
44	552	14	US-10-140-808-196	Sequence 196, App
45	552	14	US-10-121-049-196	Sequence 196, App
46	552	14	US-10-123-043-196	Sequence 196, App
47	552	14	US-10-140-740-196	Sequence 196, App
48	552	14	US-10-175-746-196	Sequence 196, App
49	552	14	US-10-176-921-196	Sequence 196, App
50	552	14	US-10-176-921-196	Sequence 196, App
51	552	14	US-10-137-865-196	Sequence 196, App
52	552	14	US-10-140-474-196	Sequence 196, App
53	552	14	US-10-142-431-196	Sequence 196, App
54	552	14	US-10-143-111-196	Sequence 196, App
55	552	14	US-10-140-002-196	Sequence 196, App
56	552	14	US-10-142-431-196	Sequence 196, App
57	552	14	US-10-123-262-196	Sequence 196, App
58	552	14	US-10-142-423-196	Sequence 196, App
59	552	14	US-10-121-050-196	Sequence 196, App
60	552	14	US-10-141-755-196	Sequence 196, App
61	552	14	US-10-143-033-196	Sequence 196, App
62	552	14	US-10-123-103-196	Sequence 196, App
63	552	14	US-10-123-236-196	Sequence 196, App
64	552	14	US-10-123-261-196	Sequence 196, App
65	552	14	US-10-140-921-196	Sequence 196, App
66	552	14	US-10-140-928-196	Sequence 196, App
67	552	14	US-10-121-045-196	Sequence 196, App
68	552	14	US-10-123-293-196	Sequence 196, App
69	552	14	US-10-123-903-196	Sequence 196, App
70	552	14	US-10-124-819-196	Sequence 196, App
71	552	14	US-10-124-822-196	Sequence 196, App
72	552	14	US-10-124-822-196	Sequence 196, App
73	552	14	US-10-160-498-196	Sequence 196, App
74	552	14	US-10-124-824-196	Sequence 196, App
75	552	14	US-10-127-825A-196	Sequence 196, App
76	552	14	US-10-127-825A-196	Sequence 196, App
77	552	14	US-10-127-835A-196	Sequence 196, App
78	552	14	US-10-127-835A-196	Sequence 196, App
79	552	14	US-10-127-901A-196	Sequence 196, App
80	552	14	US-10-128-693A-196	Sequence 196, App
81	552	14	US-10-131-813A-196	Sequence 196, App
82	552	14	US-10-131-816A-196	Sequence 196, App
83	552	14	US-10-131-823A-196	Sequence 196, App
84	552	14	US-10-131-824A-196	Sequence 196, App
85	552	14	US-10-131-830A-196	Sequence 196, App
86	552	14	US-10-131-837A-196	Sequence 196, App

87 971 29.6 552 14 US-10-137-872A-196 Sequence 196, App  
88 971 29.6 552 14 US-10-147-500-196 Sequence 196, App  
89 971 29.6 552 14 US-10-147-502-196 Sequence 196, App  
90 971 29.6 552 14 US-10-147-515-196 Sequence 196, App  
91 971 29.6 552 14 US-10-147-517-196 Sequence 196, App  
92 971 29.6 552 14 US-10-147-526-196 Sequence 196, App  
93 971 29.6 552 14 US-10-147-527-196 Sequence 196, App  
94 971 29.6 552 14 US-10-121-041-196 Sequence 196, App  
95 971 29.6 552 14 US-10-121-043-196 Sequence 196, App  
96 971 29.6 552 14 US-10-121-047-196 Sequence 196, App  
97 971 29.6 552 14 US-10-123-215-196 Sequence 196, App  
98 971 29.6 552 14 US-10-123-902-196 Sequence 196, App  
99 971 29.6 552 14 US-10-123-908-196 Sequence 196, App  
100 971 29.6 552 14 US-10-123-909-196 Sequence 196, App

## ALIGNMENTS

## RESULT 1

US-09-795-926-43  
Sequence 43, Application US/09795926  
Patent No. US20020098486A1  
GENERAL INFORMATION:  
APPLICANT: Donoho, Gregory  
APPLICANT: Halbur, Erin  
APPLICANT: Turner, C. Alexander Jr.  
APPLICANT: Friedrich, Glenn  
APPLICANT: Abulin, Alejandro  
APPLICANT: Zambrowicz, Brian  
APPLICANT: Sands, Arthur T.  
APPLICANT: Walke, D. Wade  
APPLICANT: Milgnowski, Nathaniel L.  
APPLICANT: Kleke, James Alvin  
APPLICANT: Potler, David George  
TITLE OF INVENTION: NOVEL HUMAN TRANSFERASE PROTEINS AND  
TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING THE SAME  
FILE REFERENCE: LEX-0144-USA  
CURRENT APPLICATION NUMBER: US/09/795, 926  
CURRENT FILING DATE: 2001-02-28  
PRIOR APPLICATION NUMBER: US 60/185, 920  
PRIOR FILING DATE: 2000-02-29  
PRIOR APPLICATION NUMBER: US 60/186, 558  
PRIOR FILING DATE: 2000-03-02  
PRIOR APPLICATION NUMBER: US 60/191, 849  
PRIOR FILING DATE: 2000-03-24  
NUMBER OF SEQ ID NOS: 47  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 43  
LENGTH: 603  
TYPE: PRT  
ORGANISM: homo sapiens  
US-09-795-926-43

Query Match 100.0%; Score 3278; DB 9; Length 603;  
Best Local Similarity 100.0%; Pred. No. 5e-310;  
Matches 603; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MRRERKLLQAVLALVLAALVLLPVNGLMALYRERQPDGTGGSGAAVAAPAAQSGSHSRQK 60  
DB 1 MRRERKLLQAVLALVLAALVLLPVNGLMALYRERQPDGTGGSGAAVAAPAAQSGSHSRQK 60  
QY 61 KTFPLGGQGLKMDHDEKAIIRDARVNGEQGRPYMTDAERVDQAARENGFNIIYSDK 120  
DB 61 KTFPLGGQGLKMDHDEKAIIRDARVNGEQGRPYMTDAERVDQAARENGFNIIYSDK 120  
QY 121 ISLRSPLDIRHPNCNSKRYLETLPNTSIIIPFNNEGSSLLRTVHSLVLRNSPEELVAEI 180  
DB 121 ISLRSPLDIRHPNCNSKRYLETLPNTSIIIPFNNEGSSLLRTVHSLVLRNSPEELVAEI 180  
QY 181 VLVDPSDRRLKKPLBEDYMAFPSVRIILRTKKRGLIRTRMLGASVATGDIVITFLDSHC 240  
DB 181 VLVDPSDRRLKKPLBEDYMAFPSVRIILRTKKRGLIRTRMLGASVATGDIVITFLDSHC 240

DB 181 VLVDPSDRRLKKPLBEDYMAFPSVRIILRTKKRGLIRTRMLGASVATGDIVITFLDSHC 240  
QY 241 EAVNWMLPPLDIRIARNRKTIYCPMIDVIDDDPRYETOGADMRGAFDMEYKRIPIR 300  
DB 241 EAVNWMLPPLDIRIARNRKTIYCPMIDVIDDDPRYETOGADMRGAFDMEYKRIPIR 300  
QY 301 PELQKADPSDFESPVNAGGLFAVDRKMFWEELGGYDGLBIWGGEOYEISFKVMCCGRM 360  
DB 301 PELQKADPSDFESPVNAGGLFAVDRKMFWEELGGYDGLBIWGGEOYEISFKVMCCGRM 360  
QY 361 EDIPCSRVGHIIYKRYVYKYPAGVSLARNIKRVAEVMMDYAEITTYRRRBYRLSGDV 420  
DB 361 EDIPCSRVGHIIYKRYVYKYPAGVSLARNIKRVAEVMMDYAEITTYRRRBYRLSGDV 420  
QY 421 AVQKRLSSLNCSFKFMFTKIAMDLPKFPVPEPPAAANGELRNVTGTCADTKHAGLG 480  
DB 421 AVQKRLSSLNCSFKFMFTKIAMDLPKFPVPEPPAAANGELRNVTGTCADTKHAGLG 480  
QY 481 SPLRLGCVGRGGEAANNMNVQFTFTWRREDIRPGDPQHTKKFCDAISHTSPVTLYDCHS 540  
DB 481 SPLRLGCVGRGGEAANNMNVQFTFTWRREDIRPGDPQHTKKFCDAISHTSPVTLYDCHS 540  
QY 541 MKGNQMLKTRKDKTLVHPVSGSCMDCSBDRIFMNTCNSSLTQWLFHTNSTVLEKF 600  
DB 541 MKGNQMLKTRKDKTLVHPVSGSCMDCSBDRIFMNTCNSSLTQWLFHTNSTVLEKF 600  
QY 601 NRN 603  
DB 601 NRN 603

## RESULT 2

US-10-001-851-2  
Sequence 2, Application US/10001851  
Publication No. US20020115628A1  
GENERAL INFORMATION:  
APPLICANT: MEYERS, Rachel A.  
APPLICANT: WILLIAMSON, Mark  
TITLE OF INVENTION: 47169 and 33935, No. US20020115628A1 Human Glycosyl Transferase  
TITLE OF INVENTION: Uses Thereof  
FILE REFERENCE: 10147-5601  
CURRENT APPLICATION NUMBER: US/10/001, 851  
CURRENT FILING DATE: 2001-11-20  
PRIOR APPLICATION NUMBER: US 60/249, 939  
PRIOR FILING DATE: 2000-11-20  
NUMBER OF SEQ ID NOS: 29  
SOFTWARE: Patentin Ver. 2.1  
SEQ ID NO 2  
LENGTH: 603  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-001-851-2

Query Match 100.0%; Score 3278; DB 13; Length 603;  
Best Local Similarity 100.0%; Pred. No. 5e-310;  
Matches 603; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MRRERKLLQAVLALVLAALVLLPVNGLMALYRERQPDGTGGSGAAVAAPAAQSGSHSRQK 60  
DB 1 MRRERKLLQAVLALVLAALVLLPVNGLMALYRERQPDGTGGSGAAVAAPAAQSGSHSRQK 60  
QY 61 KTFPLGGQGLKMDHDEKAIIRDARVNGEQGRPYMTDAERVDQAARENGFNIIYSDK 120  
DB 61 KTFPLGGQGLKMDHDEKAIIRDARVNGEQGRPYMTDAERVDQAARENGFNIIYSDK 120  
QY 121 ISLRSPLDIRHPNCNSKRYLETLPNTSIIIPFNNEGSSLLRTVHSLVLRNSPEELVAEI 180  
DB 121 ISLRSPLDIRHPNCNSKRYLETLPNTSIIIPFNNEGSSLLRTVHSLVLRNSPEELVAEI 180  
QY 181 VLVDPSDRRLKKPLBEDYMAFPSVRIILRTKKRGLIRTRMLGASVATGDIVITFLDSHC 240  
DB 181 VLVDPSDRRLKKPLBEDYMAFPSVRIILRTKKRGLIRTRMLGASVATGDIVITFLDSHC 240

Qy 241 EAVNMLPPLDRIARNRKTIICPMIDVIDHDDEFRYTOAGDAMRGAFDMEYKRIPI 300  
Db 241 EAVNMLPPLDRIARNRKTIICPMIDVIDHDDEFRYTOAGDAMRGAFDMEYKRIPI 300  
Qy 301 PELQADPSDFPESPVNAGGLFAVDRKMFELGSDYDGLIWSGEQYEISPKVMCCGRM 360  
Db 301 PELQADPSDFPESPVNAGGLFAVDRKMFELGSDYDGLIWSGEQYEISPKVMCCGRM 360  
Qy 361 EDIPCSRVGHIYRKVYKYPAGVSLARNLKRVAEVMMDYAEYIYORREYRHLASGDV 420  
Db 361 EDIPCSRVGHIYRKVYKYPAGVSLARNLKRVAEVMMDYAEYIYORREYRHLASGDV 420  
Qy 421 AVQKLRSSINCKSFKFMFTKIAMDLPKYPVPVPPAAANGELRNVTGICADTKHGALG 480  
Db 421 AVQKLRSSINCKSFKFMFTKIAMDLPKYPVPVPPAAANGELRNVTGICADTKHGALG 480  
Qy 481 SPLRLEGCVGRGEAANNMNOVFTFWREDIRPGDPQHTKKFCDDAISHTSPVTLYDCHS 540  
Db 481 SPLRLEGCVGRGEAANNMNOVFTFWREDIRPGDPQHTKKFCDDAISHTSPVTLYDCHS 540  
Qy 541 MKGNQLMKRYKDKTLVHPVSGSCMDCSDHRIFMNTCNPSLTOQWLFHTNSTVLEKF 600  
Db 541 MKGNQLMKRYKDKTLVHPVSGSCMDCSDHRIFMNTCNPSLTOQWLFHTNSTVLEKF 600  
Qy 601 NRN 603  
Db 601 NRN 603

## RESULT 3

US-10-364-774-43  
; Sequence 43, Application US/10364774  
; Publication No. US20030144497A1  
; GENERAL INFORMATION:  
; APPLICANT: Donoho, Gregory  
; APPLICANT: Hilbun, Erin  
; APPLICANT: Turner, C. Alexander Jr.  
; APPLICANT: Friedrich, Glenn  
; APPLICANT: Abuin, Alejandro  
; APPLICANT: Zambrowicz, Brian  
; APPLICANT: Sander, Arthur T.  
; APPLICANT: Walke, D. Wade  
; APPLICANT: Wilganowski, Nathaniel L.  
; APPLICANT: Hu, Yi  
; APPLICANT: Kiege, James Alvin  
; APPLICANT: Potter, David George  
; TITLE OF INVENTION: NOVEL HUMAN TRANSFERASE PROTEINS AND  
; FILE REFERENCE: LEX-0144-USA  
; CURRENT APPLICATION NUMBER: US/10/364,774  
; CURRENT FILING DATE: 2003-02-11  
; PRIOR APPLICATION NUMBER: US/09/795,926  
; PRIOR FILING DATE: 2001-02-28  
; PRIOR APPLICATION NUMBER: US 60/185,920  
; PRIOR FILING DATE: 2000-02-29  
; PRIOR APPLICATION NUMBER: US 60/186,558  
; PRIOR FILING DATE: 2000-03-02  
; PRIOR APPLICATION NUMBER: US 60/191,849  
; PRIOR FILING DATE: 2000-03-24  
; NUMBER OF SEQ ID NOS: 47  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 43  
; LENGTH: 603  
; TYPE: PRT  
; ORGANISM: homo sapiens  
US-10-364-774-43

Query Match 100.0%; Score 3278; DB 14; Length 603;  
Best Local Similarity 100.0%; Freq. No. 5e-310;  
Matches 603; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 MRRERKLLQAVLVLAALVLLPVNGLMALYREOPDGTGCGAATAVAPAGGSHSRQK 60  
Db 1 MRRERKLLQAVLVLAALVLLPVNGLMALYREOPDGTGCGAATAVAPAGGSHSRQK 60

Db 1 MRRERKLLQAVLVLAALVLLPVNGLMALYREOPDGTGCGAATAVAPAGGSHSRQK 60  
Qy 61 KTFPLGSGOLKQWMDHEAIRDAORVNGEQRPYMTAERDQYRENGFPIYVSDK 120  
Db 61 KTFPLGSGOLKQWMDHEAIRDAORVNGEQRPYMTAERDQYRENGFPIYVSDK 120  
Qy 121 ISLNRSLPDIRHPNCKSRILETLPTNTSIIPFNEGMSLLRTVHVLNRSPEELVAEI 180  
Db 121 ISLNRSLPDIRHPNCKSRILETLPTNTSIIPFNEGMSLLRTVHVLNRSPEELVAEI 180  
Qy 181 VLVDVDFSRHLKLPEDYNALPEPSVRLRTKREGILRTMIGASVATGVDITFLDSHC 240  
Db 181 VLVDVDFSRHLKLPEDYNALPEPSVRLRTKREGILRTMIGASVATGVDITFLDSHC 240  
Qy 241 EAVNMLPPLDRIARNRKTIICPMIDVIDHDDEFRYTOAGDAMRGAFDMEYKRIPI 300  
Db 241 EAVNMLPPLDRIARNRKTIICPMIDVIDHDDEFRYTOAGDAMRGAFDMEYKRIPI 300  
Qy 301 PELQADPSDFPESPVNAGGLFAVDRKMFELGSDYDGLIWSGEQYEISPKVMCCGRM 360  
Db 301 PELQADPSDFPESPVNAGGLFAVDRKMFELGSDYDGLIWSGEQYEISPKVMCCGRM 360  
Qy 361 EDIPCSRVGHIYRKVYKYPAGVSLARNLKRVAEVMMDYAEYIYORREYRHLASGDV 420  
Db 361 EDIPCSRVGHIYRKVYKYPAGVSLARNLKRVAEVMMDYAEYIYORREYRHLASGDV 420  
Qy 421 AVQKLRSSINCKSFKFMFTKIAMDLPKYPVPVPPAAANGELRNVTGICADTKHGALG 480  
Db 421 AVQKLRSSINCKSFKFMFTKIAMDLPKYPVPVPPAAANGELRNVTGICADTKHGALG 480  
Qy 481 SPLRLEGCVGRGEAANNMNOVFTFWREDIRPGDPQHTKKFCDDAISHTSPVTLYDCHS 540  
Db 481 SPLRLEGCVGRGEAANNMNOVFTFWREDIRPGDPQHTKKFCDDAISHTSPVTLYDCHS 540  
Qy 541 MKGNQLMKRYKDKTLVHPVSGSCMDCSDHRIFMNTCNPSLTOQWLFHTNSTVLEKF 600  
Db 541 MKGNQLMKRYKDKTLVHPVSGSCMDCSDHRIFMNTCNPSLTOQWLFHTNSTVLEKF 600  
Qy 601 NRN 603  
Db 601 NRN 603

## RESULT 4

US-09-795-926-41  
; Sequence 41, Application US/09795926  
; Patent No. US20020098486A1  
; GENERAL INFORMATION:  
; APPLICANT: Donoho, Gregory  
; APPLICANT: Hilbun, Erin  
; APPLICANT: Turner, C. Alexander Jr.  
; APPLICANT: Friedrich, Glenn  
; APPLICANT: Abuin, Alejandro  
; APPLICANT: Zambrowicz, Brian  
; APPLICANT: Sander, Arthur T.  
; APPLICANT: Walke, D. Wade  
; APPLICANT: Wilganowski, Nathaniel L.  
; APPLICANT: Hu, Yi  
; APPLICANT: Kiege, James Alvin  
; APPLICANT: Potter, David George  
; TITLE OF INVENTION: NOVEL HUMAN TRANSFERASE PROTEINS AND  
; FILE REFERENCE: LEX-0144-USA  
; CURRENT APPLICATION NUMBER: US/09/795,926  
; CURRENT FILING DATE: 2001-02-28  
; PRIOR APPLICATION NUMBER: US 60/185,920  
; PRIOR FILING DATE: 2000-02-29  
; PRIOR APPLICATION NUMBER: US 60/186,558  
; PRIOR FILING DATE: 2000-03-02  
; PRIOR APPLICATION NUMBER: US 60/191,849  
; PRIOR FILING DATE: 2000-03-24  
; NUMBER OF SEQ ID NOS: 47  
; SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 41  
LENGTH: 631  
TYPE: PR  
ORGANISM: homo sapiens  
US-09-795-926-41

Query Match 98.9%; Score 3242.5; DB 9; Length 631;  
Best Local Similarity 95.4%; Pred. No. 1.5e-306;  
Matches 601; Conservative 0; Mismatches 0; Indels 29; Gaps 1;

```
QY 1 MRRREKLLQVALVLAVALVLLPNVGLMALYRERPDGTPGSGAVALPAAGQSHSRQK 60
DB 1 MRRREKLLQVALVLAVALVLLPNVGLMALYRERPDGTPGSGAVALPAAGQSHSRQK 60
QY 61 KTFPLGGQQLKMDHDEKAIIRDARVNGEGRPYPTDAERVDQAYRENGFNIVSDK 120
DB 61 KTFPLGGQQLKMDHDEKAIIRDARVNGEGRPYPTDAERVDQAYRENGFNIVSDK 120
QY 121 ISLRSLEPDIRHPNCNKRILETLPNTSIIIPFNHGMSSLLRTVHSLNRSPELVAEI 180
DB 121 ISLRSLEPDIRHPNCNKRILETLPNTSIIIPFNHGMSSLLRTVHSLNRSPELVAEI 180
QY 181 VLVDPSDRREHLKKPLEDYMALFPSVRLRTKKREGILRTMLGASVATGDTVITFLDSHC 240
DB 181 VLVDPSDRREHLKKPLEDYMALFPSVRLRTKKREGILRTMLGASVATGDTVITFLDSHC 240
QY 241 EANNWMLPPLDRLARRKRTIVCPMIDVIDHDDPRYETQAGDAMRGAFDMEMYKRIPI 300
DB 241 EANNWMLPPLDRLARRKRTIVCPMIDVIDHDDPRYETQAGDAMRGAFDMEMYKRIPI 300
QY 301 PELQKADPSDFESPVMAGSLFAVDRKMFELGSDYDGLBETWGEQYEISFGKLHMLPRL 352
DB 301 PELQKADPSDFESPVMAGSLFAVDRKMFELGSDYDGLBETWGEQYEISFGKLHMLPRL 352
QY 353 -----VMMCGRMEDIPCSRGHITRYKYVPYKVPAGVSLARRLK 391
DB 353 -----VMMCGRMEDIPCSRGHITRYKYVPYKVPAGVSLARRLK 391
QY 392 RVAEVMWDEYAEIYQRRPEYRHLASGDVAVQKLRSLNCKSPKMTKIAMLPKPY 451
DB 392 RVAEVMWDEYAEIYQRRPEYRHLASGDVAVQKLRSLNCKSPKMTKIAMLPKPY 451
QY 452 PVEPPAAWGEIRNVGTGLCADTGHGALGSLPLREGCVRGGEAAMNNQVFTFTWREDI 511
DB 452 PVEPPAAWGEIRNVGTGLCADTGHGALGSLPLREGCVRGGEAAMNNQVFTFTWREDI 511
QY 512 RPDGPQHTKKCFDAISHTSPVTLVDCHSMKGNQIMKRYKDTLYHPVSGSCMDCSESDH 571
DB 512 RPDGPQHTKKCFDAISHTSPVTLVDCHSMKGNQIMKRYKDTLYHPVSGSCMDCSESDH 571
QY 572 RIFMNTCNPSLSLTOQMLFEHTNSTVLEKEN 601
DB 572 RIFMNTCNPSLSLTOQMLFEHTNSTVLEKEN 601
QY 601 RIFMNTCNPSLSLTOQMLFEHTNSTVLEKEN 630
DB 601 RIFMNTCNPSLSLTOQMLFEHTNSTVLEKEN 630
```

## RESULT 5

US-10-364-774-41  
Sequence 41, Application US/10364774  
Publication No. US20030144497A1  
GENERAL INFORMATION:  
APPLICANT: Donoho, Gregory  
APPLICANT: Hilbun, Erin  
APPLICANT: Turner, C. Alexander Jr.  
APPLICANT: Friedlich, Glenn  
APPLICANT: Abutin, Alejandro  
APPLICANT: Zambrowicz, Brian  
APPLICANT: Sands, Arthur T.  
APPLICANT: Walke, D. Wade  
APPLICANT: Milganowski, Nathaniel L.  
APPLICANT: Hu, Yi  
APPLICANT: Kieke, James Alvin  
APPLICANT: Potter, David George  
TITLE OF INVENTION: NOVEL HUMAN TRANSFERASE PROTEINS AND

TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING THE SAME  
FILE REFERENCE: LEX-0144-USA  
CURRENT APPLICATION NUMBER: US/10/364, 774  
CURRENT FILING DATE: 2003-02-11  
PRIOR APPLICATION NUMBER: US/09/795, 926  
PRIOR FILING DATE: 2001-02-28  
PRIOR APPLICATION NUMBER: US 60/185, 920  
PRIOR FILING DATE: 2000-02-29  
PRIOR APPLICATION NUMBER: US 60/186, 558  
PRIOR FILING DATE: 2000-03-02  
PRIOR APPLICATION NUMBER: US 60/191, 849  
NUMBER OF SEQ ID NOS: 47  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 41  
LENGTH: 631  
TYPE: PR  
ORGANISM: homo sapiens  
US-10-364-774-41

Query Match 98.9%; Score 3242.5; DB 14; Length 631;  
Best Local Similarity 95.4%; Pred. No. 1.5e-306;  
Matches 601; Conservative 0; Mismatches 0; Indels 29; Gaps 1;

```
QY 1 MRRREKLLQVALVLAVALVLLPNVGLMALYRERPDGTPGSGAVALPAAGQSHSRQK 60
DB 1 MRRREKLLQVALVLAVALVLLPNVGLMALYRERPDGTPGSGAVALPAAGQSHSRQK 60
QY 61 KTFPLGGQQLKMDHDEKAIIRDARVNGEGRPYPTDAERVDQAYRENGFNIVSDK 120
DB 61 KTFPLGGQQLKMDHDEKAIIRDARVNGEGRPYPTDAERVDQAYRENGFNIVSDK 120
QY 121 ISLRSLEPDIRHPNCNKRILETLPNTSIIIPFNHGMSSLLRTVHSLNRSPELVAEI 180
DB 121 ISLRSLEPDIRHPNCNKRILETLPNTSIIIPFNHGMSSLLRTVHSLNRSPELVAEI 180
QY 181 VLVDPSDRREHLKKPLEDYMALFPSVRLRTKKREGILRTMLGASVATGDTVITFLDSHC 240
DB 181 VLVDPSDRREHLKKPLEDYMALFPSVRLRTKKREGILRTMLGASVATGDTVITFLDSHC 240
QY 241 EANNWMLPPLDRLARRKRTIVCPMIDVIDHDDPRYETQAGDAMRGAFDMEMYKRIPI 300
DB 241 EANNWMLPPLDRLARRKRTIVCPMIDVIDHDDPRYETQAGDAMRGAFDMEMYKRIPI 300
QY 301 PELQKADPSDFESPVMAGSLFAVDRKMFELGSDYDGLBETWGEQYEISFGKLHMLPRL 352
DB 301 PELQKADPSDFESPVMAGSLFAVDRKMFELGSDYDGLBETWGEQYEISFGKLHMLPRL 352
QY 353 -----VMMCGRMEDIPCSRGHITRYKYVPYKVPAGVSLARRLK 391
DB 353 -----VMMCGRMEDIPCSRGHITRYKYVPYKVPAGVSLARRLK 391
QY 392 RVAEVMWDEYAEIYQRRPEYRHLASGDVAVQKLRSLNCKSPKMTKIAMLPKPY 451
DB 392 RVAEVMWDEYAEIYQRRPEYRHLASGDVAVQKLRSLNCKSPKMTKIAMLPKPY 451
QY 452 PVEPPAAWGEIRNVGTGLCADTGHGALGSLPLREGCVRGGEAAMNNQVFTFTWREDI 511
DB 452 PVEPPAAWGEIRNVGTGLCADTGHGALGSLPLREGCVRGGEAAMNNQVFTFTWREDI 511
QY 512 RPDGPQHTKKCFDAISHTSPVTLVDCHSMKGNQIMKRYKDTLYHPVSGSCMDCSESDH 571
DB 512 RPDGPQHTKKCFDAISHTSPVTLVDCHSMKGNQIMKRYKDTLYHPVSGSCMDCSESDH 571
QY 572 RIFMNTCNPSLSLTOQMLFEHTNSTVLEKEN 601
DB 572 RIFMNTCNPSLSLTOQMLFEHTNSTVLEKEN 601
QY 601 RIFMNTCNPSLSLTOQMLFEHTNSTVLEKEN 630
DB 601 RIFMNTCNPSLSLTOQMLFEHTNSTVLEKEN 630
```

## RESULT 6

US-09-795-926-31  
Sequence 31, Application US/09795926  
Patent No. US2002009486A1



Qy 578 CNPSSLTQOMLFEHTNSTVLEKFNRN 603  
Db 481 CNPSSLTQOMLFEHTNSTVLEKFNRN 506

RESULT 8  
US-09-795-926-29

Sequence 29, Application US/09795926  
Patent No. US20020098486A1  
GENERAL INFORMATION:  
APPLICANT: Donoho, Gregory  
APPLICANT: Hilbun, Erin  
APPLICANT: Turner, C. Alexander Jr.  
APPLICANT: Friedrich, Glenn  
APPLICANT: Abulin, Alejandro  
APPLICANT: Zambrowicz, Brian  
APPLICANT: Sands, Arthur T.  
APPLICANT: Walke, D. Wade  
APPLICANT: Milganowski, Nathaniel L.  
APPLICANT: Hu, Yi  
APPLICANT: Kleke, James Alvin  
APPLICANT: Potter, David George  
TITLE OF INVENTION: NOVEL HUMAN TRANSFERASE PROTEINS AND  
FILE REFERENCE: LEX-0144-USA  
CURRENT FILING DATE: 2001-02-28  
PRIOR FILING DATE: 2000-02-29  
PRIOR APPLICATION NUMBER: US 60/185,920  
PRIOR FILING DATE: 2000-02-29  
PRIOR APPLICATION NUMBER: US 60/186,558  
PRIOR FILING DATE: 2000-03-02  
PRIOR APPLICATION NUMBER: US 60/191,849  
PRIOR FILING DATE: 2000-03-24  
NUMBER OF SEQ ID NOS: 47  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 29  
LENGTH: 535  
TYPE: PRT  
ORGANISM: homo sapiens  
US-09-795-926-29

Query Match 83.8%; Score 2746.5; DB 9; Length 535;  
Best Local Similarity 94.6%; Pred. No. 2.8e-258;  
Matches 506; Conservative 0; Mismatches 0; Indels 29; Gaps 1;

Qy 98 MTDARVDQARENGFNIVYSDKISLNRSLPDIRHPCNSKRYLETLPNTSIIIPFNEG 157  
Db 1 MTDARVDQARENGFNIVYSDKISLNRSLPDIRHPCNSKRYLETLPNTSIIIPFNEG 60  
Qy 158 WSSLIRTVHSLVNSPPELVAEIYLVDDFSDREHLKPLEDYMALPFSVRILRTKKEGL 217  
Db 61 WSSLIRTVHSLVNSPPELVAEIYLVDDFSDREHLKPLEDYMALPFSVRILRTKKEGL 120  
Qy 218 IRTMLGASVATGVITFLDSHCANVWMLPPLDRIARNKTIVCPMIDVIDHDDFRYE 277  
Db 121 IRTMLGASVATGVITFLDSHCANVWMLPPLDRIARNKTIVCPMIDVIDHDDFRYE 180  
Qy 278 TQADNARGAFDWMYTKRIPPELOKADPSDFESPVAAGLFAVDRKWFELGSDP 337  
Db 181 TQADNARGAFDWMYTKRIPPELOKADPSDFESPVAAGLFAVDRKWFELGSDP 240  
Qy 338 GLEIHWGEOYEISFK-----VMMCGRMEDIPCSRV 368  
Db 241 GLEIHWGEOYEISFKGLHMLPRLVSNWPOAVFLPRAPNMLALOVMMCGRMEDIPCSRV 300  
Qy 369 GHIYRKVPYKVPAGVSLARNLKRVAEVMDEVAEYIYQRRPEYRHLISAGDAVQKRLS 428  
Db 301 GHIYRKVPYKVPAGVSLARNLKRVAEVMDEVAEYIYQRRPEYRHLISAGDAVQKRLS 360  
Qy 429 SLNCKSFEMWTKIAMDLPKFYPPVEPPAAWGEIRNVTGLCADTKHAGLSPLRLGCG 488  
Db 361 SLNCKSFEMWTKIAMDLPKFYPPVEPPAAWGEIRNVTGLCADTKHAGLSPLRLGCG 420

Qy 489 VRGRGAAMNMNQVFTFTWRBEDIKPGDPQHTKKCFDAISHTSEVTLYDCHSKGNQJWK 548  
Db 421 VRGRGAAMNMNQVFTFTWRBEDIKPGDPQHTKKCFDAISHTSEVTLYDCHSKGNQJWK 480  
Qy 549 YRKDKTLVHPVSGSCMDCSSDRIEMNTCNPSSLTQOMLFEHTNSTVLEKFNRN 603  
Db 481 YRKDKTLVHPVSGSCMDCSSDRIEMNTCNPSSLTQOMLFEHTNSTVLEKFNRN 535

RESULT 9  
US-10-364-774-29

Sequence 29, Application US/10364774  
Publication No. US20030144497A1  
GENERAL INFORMATION:  
APPLICANT: Donoho, Gregory  
APPLICANT: Hilbun, Erin  
APPLICANT: Turner, C. Alexander Jr.  
APPLICANT: Friedrich, Glenn  
APPLICANT: Abulin, Alejandro  
APPLICANT: Zambrowicz, Brian  
APPLICANT: Sands, Arthur T.  
APPLICANT: Walke, D. Wade  
APPLICANT: Milganowski, Nathaniel L.  
APPLICANT: Hu, Yi  
APPLICANT: Kleke, James Alvin  
APPLICANT: Potter, David George  
TITLE OF INVENTION: NOVEL HUMAN TRANSFERASE PROTEINS AND  
FILE REFERENCE: LEX-0144-USA  
CURRENT FILING DATE: 2003-02-11  
PRIOR FILING DATE: 2001-02-28  
PRIOR APPLICATION NUMBER: US/10/364,774  
PRIOR FILING DATE: 2000-02-29  
PRIOR APPLICATION NUMBER: US 60/185,920  
PRIOR FILING DATE: 2000-02-29  
PRIOR APPLICATION NUMBER: US 60/186,558  
PRIOR FILING DATE: 2000-03-02  
PRIOR APPLICATION NUMBER: US 60/191,849  
PRIOR FILING DATE: 2000-03-24  
NUMBER OF SEQ ID NOS: 47  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 29  
LENGTH: 535  
TYPE: PRT  
ORGANISM: homo sapiens  
US-10-364-774-29

Query Match 83.8%; Score 2746.5; DB 14; Length 535;  
Best Local Similarity 94.6%; Pred. No. 2.8e-258;  
Matches 506; Conservative 0; Mismatches 0; Indels 29; Gaps 1;

Qy 98 MTDARVDQARENGFNIVYSDKISLNRSLPDIRHPCNSKRYLETLPNTSIIIPFNEG 157  
Db 1 MTDARVDQARENGFNIVYSDKISLNRSLPDIRHPCNSKRYLETLPNTSIIIPFNEG 60  
Qy 158 WSSLIRTVHSLVNSPPELVAEIYLVDDFSDREHLKPLEDYMALPFSVRILRTKKEGL 217  
Db 61 WSSLIRTVHSLVNSPPELVAEIYLVDDFSDREHLKPLEDYMALPFSVRILRTKKEGL 120  
Qy 218 IRTMLGASVATGVITFLDSHCANVWMLPPLDRIARNKTIVCPMIDVIDHDDFRYE 277  
Db 121 IRTMLGASVATGVITFLDSHCANVWMLPPLDRIARNKTIVCPMIDVIDHDDFRYE 180  
Qy 278 TQADNARGAFDWMYTKRIPPELOKADPSDFESPVAAGLFAVDRKWFELGSDP 337  
Db 181 TQADNARGAFDWMYTKRIPPELOKADPSDFESPVAAGLFAVDRKWFELGSDP 240  
Qy 338 GLEIHWGEOYEISFK-----VMMCGRMEDIPCSRV 368  
Db 241 GLEIHWGEOYEISFKGLHMLPRLVSNWPOAVFLPRAPNMLALOVMMCGRMEDIPCSRV 300  
Qy 369 GHIYRKVPYKVPAGVSLARNLKRVAEVMDEVAEYIYQRRPEYRHLISAGDAVQKRLS 428

Db 301 GHIYKAVPVKVPAGVSLARLKVAVMDEYAYIYORPEPRHLSAGDVAVQKLR 360  
Qy 429 SLNCKSFEMTNTKIAMDLPKPYPEVPEPPAAAGSIRNVGTGLCADYTKHAGLSPLREGC 488  
Db 361 SLNCKSFEMTNTKIAMDLPKPYPEVPEPPAAAGSIRNVGTGLCADYTKHAGLSPLREGC 420  
Qy 489 VRGGEAANNMNOVFTTWEDIRPGDPORHTKKECFPAISHTSPVTLYDCSHKMGQNLWK 548  
Db 421 VRGGEAANNMNOVFTTWEDIRPGDPORHTKKECFPAISHTSPVTLYDCSHKMGQNLWK 480  
Qy 549 YRKDKTYHPVSGSCMDCSESDHRIFMNTCNPSLSLTQOMLPEHNTSTVLEKFNEN 603  
Db 481 YRKDKTYHPVSGSCMDCSESDHRIFMNTCNPSLSLTQOMLPEHNTSTVLEKFNEN 535  
RESULT 10  
US-09-795-926-39  
; Sequence 39, Application US/09795926  
; Patent No. US20020098486A1  
; GENERAL INFORMATION:  
; APPLICANT: Donoho, Gregory  
; APPLICANT: Hilbun, Erin  
; APPLICANT: Turner, C. Alexander Jr.  
; APPLICANT: Friedrich, Glenn  
; APPLICANT: Abuin, Alejandro  
; APPLICANT: Zambrowicz, Brian  
; APPLICANT: Sander, Arthur T.  
; APPLICANT: Walke, D. Wade  
; APPLICANT: Milganowski, Nathaniel L.  
; APPLICANT: Hu, Yi  
; APPLICANT: Kieck, James Alvin  
; APPLICANT: Potter, David George  
; TITLE OF INVENTION: NOVEL HUMAN TRANSFERASE PROTEINS AND  
; FILE REFERENCE: LEX-0144-USA  
; CURRENT APPLICATION NUMBER: US/09/795,926  
; PRIOR FILING DATE: 2001-02-28  
; PRIOR APPLICATION NUMBER: US 60/185,920  
; PRIOR FILING DATE: 2000-02-29  
; PRIOR APPLICATION NUMBER: US 60/186,558  
; PRIOR FILING DATE: 2000-03-02  
; PRIOR APPLICATION NUMBER: US 60/191,849  
; PRIOR FILING DATE: 2000-03-24  
; NUMBER OF SEQ ID NOS: 47  
; SOFTWARE: FASTSEQ for Windows Version 4.0  
; SEQ ID NO 39  
; LENGTH: 366  
; TYPE: PRT  
; ORGANISM: homo sapiens  
US-09-795-926-39  
Query Match 57.3%; Score 1877; DB 9; Length 366;  
Best Local Similarity 100.0%; Pred. No. 9,5e-174;  
Matches 353; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 241 EAVNMVLPPLDDRJAARRKTIIVCPMIDVIDHDDFRYETQAGDARGAPEWEMTYKRIPIP 300  
Qy 301 PELQKADPSDFESPVMAAGLPAVDRKWFELGSDYDGLIEWGGEQYEISFKV 353  
Db 301 PELQKADPSDFESPVMAAGLPAVDRKWFELGSDYDGLIEWGGEQYEISFKV 353  
RESULT 11  
US-10-364-774-39  
; Sequence 39, Application US/10364774  
; Publication No. US20030144497A1  
; GENERAL INFORMATION:  
; APPLICANT: Donoho, Gregory  
; APPLICANT: Hilbun, Erin  
; APPLICANT: Turner, C. Alexander Jr.  
; APPLICANT: Friedrich, Glenn  
; APPLICANT: Abuin, Alejandro  
; APPLICANT: Zambrowicz, Brian  
; APPLICANT: Sander, Arthur T.  
; APPLICANT: Walke, D. Wade  
; APPLICANT: Milganowski, Nathaniel L.  
; APPLICANT: Hu, Yi  
; APPLICANT: Kieck, James Alvin  
; APPLICANT: Potter, David George  
; TITLE OF INVENTION: NOVEL HUMAN TRANSFERASE PROTEINS AND  
; FILE REFERENCE: LEX-0144-USA  
; CURRENT APPLICATION NUMBER: US/10/364,774  
; PRIOR FILING DATE: 2003-02-11  
; PRIOR APPLICATION NUMBER: US/09/795,926  
; PRIOR FILING DATE: 2001-02-28  
; PRIOR APPLICATION NUMBER: US 60/185,920  
; PRIOR FILING DATE: 2000-02-29  
; PRIOR APPLICATION NUMBER: US 60/186,558  
; PRIOR FILING DATE: 2000-03-02  
; PRIOR APPLICATION NUMBER: US 60/191,849  
; PRIOR FILING DATE: 2000-03-24  
; NUMBER OF SEQ ID NOS: 47  
; SOFTWARE: FASTSEQ for Windows Version 4.0  
; SEQ ID NO 39  
; LENGTH: 366  
; TYPE: PRT  
; ORGANISM: homo sapiens  
US-10-364-774-39  
Query Match 57.3%; Score 1877; DB 14; Length 366;  
Best Local Similarity 100.0%; Pred. No. 9,5e-174;  
Matches 353; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 12  
US-10-292-896-3  
Sequence 3, Application US/10292896  
Publication No. US20030186850A1  
GENERAL INFORMATION:  
APPLICANT: HASSAN, Helle  
APPLICANT: REIS, Celso A.  
APPLICANT: BENNETT, Eric P.  
APPLICANT: CLAUSEN, Henrik  
TITLE OF INVENTION: METHODS TO IDENTIFY AGENTS MODULATING FUNCTIONS OF POLYPEPTIDE GA  
TITLE OF INVENTION: TRANSFERASIS, PHARMACEUTICAL COMPOSITIONS COMPRISING SUCH AGENTS  
FILE REFERENCE: 4305/1H154-US3  
CURRENT APPLICATION NUMBER: US/10/292,896  
PRIOR FILING DATE: 2003-03-17  
PRIOR APPLICATION NUMBER: US 60/425,204  
PRIOR FILING DATE: 2002-11-08  
PRIOR APPLICATION NUMBER: PCT/DK01/00328  
PRIOR FILING DATE: 2001-05-10  
PRIOR APPLICATION NUMBER: US 60/203,331  
PRIOR FILING DATE: 2000-05-11  
NUMBER OF SEQ ID NOS: 130  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 3  
LENGTH: 276  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-292-896-3

Query Match 42.5%; Score 1392; DB 14; Length 276;  
Best Local Similarity 97.3%; Pred. No. 1.2e-126;  
Matches 250; Conservative 3; Mismatches 2; Indels 2; Gaps 1;

Qy 349 ISFKVMGGMEDIPCSRVGHITRYKYVPYKVPAGVSLA--RNLKRAVEVMMDRYAEITY 406  
Db 20 LAMVVMGGMEDIPCSRVGHITRYKYVPYKVPAGVSLA--RNLKRAVEVMMDRYAEITY 79  
Qy 407 ORREYRLHLSAGDVAVOCKRLSSLNCKSGFKMFMPTIADLPKPYPPVPPPAANGELINV 466  
Db 80 ORREYRLHLSAGDVAVOCKRLSSLNCKSGFKMFMPTIADLPKPYPPVPPPAANGELINV 139  
Qy 467 GTGCAADTKGALSPRLBEGCVRGREGAANNMVFPTFRREDIRPGDPQHTKKFCFDA 526  
Db 140 GTGCAADTKGALSPRLBEGCVRGREGAANNMVFPTFRREDIRPGDPQHTKKFCFDA 199  
Qy 527 ISHTSPVTLVYCHSKMGQMLKRYKDKTLVHPVSSCMDCESDHRIEMTICNPSLTQQ 586  
Db 200 ISHTSPVTLVYCHSKMGQMLKRYKDKTLVHPVSSCMDCESDHRIEMTICNPSLTQQ 259  
Qy 587 WLFEHNTSTVLEKFNRN 603  
Db 260 WLFEHNTSTVLEKFNRN 276

RESULT 13  
US-09-795-926-27  
Sequence 27, Application US/09795926  
Patent No. US20020098486A1  
GENERAL INFORMATION:  
APPLICANT: Donoho, Gregory  
APPLICANT: Hilbun, Erin  
APPLICANT: Turner, C. Alexander Jr.  
APPLICANT: Friedrich, Glenn  
APPLICANT: Abuin, Alejandro  
APPLICANT: Zambrowicz, Brian  
APPLICANT: Sands, Arthur T.  
APPLICANT: Walke, D. Wade  
APPLICANT: Wilganowski, Nathaniel L.  
APPLICANT: Hu, Yi  
APPLICANT: Kieke, James Alvin  
APPLICANT: Potter, David George

TITLE OF INVENTION: NOVEL HUMAN TRANSFERASE PROTEINS AND  
TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING THE SAME  
FILE REFERENCE: LEX-0144-USA  
CURRENT APPLICATION NUMBER: US/09/795,926  
CURRENT FILING DATE: 2001-02-28  
PRIOR APPLICATION NUMBER: US 60/185,920  
PRIOR FILING DATE: 2000-02-29  
PRIOR APPLICATION NUMBER: US 60/186,558  
PRIOR FILING DATE: 2000-03-02  
PRIOR APPLICATION NUMBER: US 60/191,849  
PRIOR FILING DATE: 2000-03-24  
NUMBER OF SEQ ID NOS: 47  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 27  
LENGTH: 269  
TYPE: PRT  
ORGANISM: homo sapiens  
US-09-795-926-27

Query Match 41.8%; Score 1370; DB 9; Length 269;  
Best Local Similarity 100.0%; Pred. No. 1.6e-124;  
Matches 256; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 98 MTDARVDOAYRENGFNITVSDKISLRSLPDIRHPCNSKRYLETIPNTSIIIPFNEG 157  
Db 1 MTDARVDOAYRENGFNITVSDKISLRSLPDIRHPCNSKRYLETIPNTSIIIPFNEG 60  
Qy 158 WSLIRTVHSVYLSNPSPELVAEIVLVDFSDREHLKKPLDYMLPFSVAILRTKKEGL 217  
Db 61 WSLIRTVHSVYLSNPSPELVAEIVLVDFSDREHLKKPLDYMLPFSVAILRTKKEGL 120  
Qy 218 IRTMLGASVATGVDITFLDSHCANVNLPLLDRIARRKTIIVCEMIDVIDHDDRYE 277  
Db 121 IRTMLGASVATGVDITFLDSHCANVNLPLLDRIARRKTIIVCEMIDVIDHDDRYE 180  
Qy 278 TQAGDARGAFDMYTKRIPPELOKADPSDFESPVAAGLFAVDKRMFELGYDP 337  
Db 181 TQAGDARGAFDMYTKRIPPELOKADPSDFESPVAAGLFAVDKRMFELGYDP 240  
Qy 338 GLEIWMGEQYEISPKV 353  
Db 241 GLEIWMGEQYEISPKV 256

RESULT 14  
US-10-364-774-27  
Sequence 27, Application US/10364774  
Publication No. US20030144497A1  
GENERAL INFORMATION:  
APPLICANT: Donoho, Gregory  
APPLICANT: Hilbun, Erin  
APPLICANT: Turner, C. Alexander Jr.  
APPLICANT: Friedrich, Glenn  
APPLICANT: Abuin, Alejandro  
APPLICANT: Zambrowicz, Brian  
APPLICANT: Sands, Arthur T.  
APPLICANT: Walke, D. Wade  
APPLICANT: Wilganowski, Nathaniel L.  
APPLICANT: Hu, Yi  
APPLICANT: Kieke, James Alvin  
APPLICANT: Potter, David George  
TITLE OF INVENTION: NOVEL HUMAN TRANSFERASE PROTEINS AND  
TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING THE SAME  
FILE REFERENCE: LEX-0144-USA  
CURRENT APPLICATION NUMBER: US/10/364,774  
CURRENT FILING DATE: 2003-02-11  
PRIOR APPLICATION NUMBER: US/09/795,926  
PRIOR FILING DATE: 2001-02-28  
PRIOR APPLICATION NUMBER: US 60/185,920  
PRIOR FILING DATE: 2000-02-29  
PRIOR APPLICATION NUMBER: US 60/186,558  
PRIOR FILING DATE: 2000-03-02  
PRIOR APPLICATION NUMBER: US 60/191,849

PRIOR FILING DATE: 2000-03-24  
NUMBER OF SEQ ID NOS: 47  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 27  
LENGTH: 269  
TYPE: PRT  
ORGANISM: homo sapiens  
US-10-364-774-27

Query Match 41.8%; Score 1370; DB 14; Length 269;  
Best Local Similarity 100.0%; Pred. No. 1.6e-124;  
Matches 256; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 98 MTDARVDAQAREGPNIVYSDKISLNRLPDIRHPNCKRYLETLPNTSIIIPFNREG 157  
DB 1 MTDARVDAQAREGPNIVYSDKISLNRLPDIRHPNCKRYLETLPNTSIIIPFNREG 60  
QY 158 WSSILRTVHSLVNSPPELVAEIVLVDDFSDREHLKKPLEDYMALPPSVRLRTKREGI 217  
DB 61 WSSILRTVHSLVNSPPELVAEIVLVDDFSDREHLKKPLEDYMALPPSVRLRTKREGI 120  
QY 218 IRTMLGASVATGVITFLDSHCANVMPLPDLRIARNKTIVCPMIDVIDHDDFRYE 277  
DB 121 IRTMLGASVATGVITFLDSHCANVMPLPDLRIARNKTIVCPMIDVIDHDDFRYE 180  
QY 278 TQAGDARGAPDMEMYYKRIPPELOKADSDPESFVMAAGLPAVDRKMFELGYPD 337  
DB 181 TQAGDARGAPDMEMYYKRIPPELOKADSDPESFVMAAGLPAVDRKMFELGYPD 240  
QY 338 GLBIWGEQYEISFKV 353  
DB 241 GLBIWGEQYEISFKV 256

RESULT 15  
US-09-795-926-35  
Sequence 35, Application US/09795926  
Patent No. US2002009846A1  
GENERAL INFORMATION:  
APPLICANT: Donoho, Gregory  
APPLICANT: Hilbun, Erin  
APPLICANT: Turner, C. Alexander Jr.  
APPLICANT: Friedrich, Glenn  
APPLICANT: Abuin, Alejandro  
APPLICANT: Zamdrowicz, Brian  
APPLICANT: Sands, Arthur T.  
APPLICANT: Walke, D. Wade  
APPLICANT: Milganowski, Nathaniel L.  
APPLICANT: Hu, Yi  
APPLICANT: Kieke, James Alvin  
APPLICANT: Potter, David George  
TITLE OF INVENTION: NOVEL HUMAN TRANSFERASE PROTEINS AND  
FILE REFERENCE: LEX-0144-USA  
CURRENT APPLICATION NUMBER: US/09/795,926  
PRIOR FILING DATE: 2001-02-28  
PRIOR APPLICATION NUMBER: US 60/185,920  
PRIOR FILING DATE: 2000-02-29  
PRIOR APPLICATION NUMBER: US 60/186,558  
PRIOR FILING DATE: 2000-03-02  
PRIOR APPLICATION NUMBER: US 60/191,849  
PRIOR FILING DATE: 2000-03-24  
NUMBER OF SEQ ID NOS: 47  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 35  
LENGTH: 321  
TYPE: PRT  
ORGANISM: homo sapiens  
US-09-795-926-35

Query Match 40.0%; Score 1311; DB 9; Length 321;  
Best Local Similarity 99.2%; Pred. No. 1.2e-118;  
Matches 251; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 MRRKREKLLQAVLVLAALVTLPLVNGLMALYREROPDGTGGSGAAYAPAGGSHSRK 60  
DB 1 MRRKREKLLQAVLVLAALVTLPLVNGLMALYREROPDGTGGSGAAYAPAGGSHSRK 60  
QY 61 KTFPLDGGQKLKQWHDKEAIRDAQRYNGEQRPPYMTDAERYDAQYRENGFNIVYSDK 120  
DB 61 KTFPLDGGQKLKQWHDKEAIRDAQRYNGEQRPPYMTDAERYDAQYRENGFNIVYSDK 120  
QY 121 ISLNRSLPDIRHPNCKRYLETLPNTSIIIPFNREGSSLLRTVHSLVNSPPELVAEI 180  
DB 121 ISLNRSLPDIRHPNCKRYLETLPNTSIIIPFNREGSSLLRTVHSLVNSPPELVAEI 180  
QY 181 VLVDPSDRHLKKPLEDYMALPPSVRLRTKREGIIRRMGASVATGVITFLDSHC 240  
DB 181 VLVDPSDRHLKKPLEDYMALPPSVRLRTKREGIIRRMGASVATGVITFLDSHC 240  
QY 241 EANYVMPLPDLDR 253  
DB 241 EANYVMPLPDLDR 253

RESULT 16  
US-10-364-774-35  
Sequence 35, Application US/10364774  
Publication No. US20030144497A1  
GENERAL INFORMATION:  
APPLICANT: Donoho, Gregory  
APPLICANT: Hilbun, Erin  
APPLICANT: Turner, C. Alexander Jr.  
APPLICANT: Friedrich, Glenn  
APPLICANT: Abuin, Alejandro  
APPLICANT: Zamdrowicz, Brian  
APPLICANT: Sands, Arthur T.  
APPLICANT: Walke, D. Wade  
APPLICANT: Milganowski, Nathaniel L.  
APPLICANT: Hu, Yi  
APPLICANT: Kieke, James Alvin  
APPLICANT: Potter, David George  
TITLE OF INVENTION: NOVEL HUMAN TRANSFERASE PROTEINS AND  
FILE REFERENCE: LEX-0144-USA  
CURRENT APPLICATION NUMBER: US/10/364,774  
PRIOR FILING DATE: 2003-02-11  
PRIOR APPLICATION NUMBER: US/09/795,926  
PRIOR FILING DATE: 2001-02-28  
PRIOR APPLICATION NUMBER: US 60/185,920  
PRIOR FILING DATE: 2000-02-29  
PRIOR APPLICATION NUMBER: US 60/186,558  
PRIOR FILING DATE: 2000-03-02  
PRIOR APPLICATION NUMBER: US 60/191,849  
PRIOR FILING DATE: 2000-03-24  
NUMBER OF SEQ ID NOS: 47  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 35  
LENGTH: 321  
TYPE: PRT  
ORGANISM: homo sapiens  
US-10-364-774-35

Query Match 40.0%; Score 1311; DB 14; Length 321;  
Best Local Similarity 99.2%; Pred. No. 1.2e-118;  
Matches 251; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 MRRKREKLLQAVLVLAALVTLPLVNGLMALYREROPDGTGGSGAAYAPAGGSHSRK 60  
DB 1 MRRKREKLLQAVLVLAALVTLPLVNGLMALYREROPDGTGGSGAAYAPAGGSHSRK 60  
QY 61 KTFPLDGGQKLKQWHDKEAIRDAQRYNGEQRPPYMTDAERYDAQYRENGFNIVYSDK 120  
DB 61 KTFPLDGGQKLKQWHDKEAIRDAQRYNGEQRPPYMTDAERYDAQYRENGFNIVYSDK 120  
QY 121 ISLNRSLPDIRHPNCKRYLETLPNTSIIIPFNREGSSLLRTVHSLVNSPPELVAEI 180

Db 121 ISLRSJLPDHRHPCNSKRYLETLPNTSIIIPFHNEGSSLLRTVHSLVNSRPPELVAEI 180  
Qy 181 VLVDPSREHLKKPLEDYMALPESVRILRTKKEGRLRTMTLGSVATGVTIFLDSHC 240  
Db 181 VLVDPSREHLKKPLEDYMALPESVRILRTKKEGRLRTMTLGSVATGVTIFLDSHC 240  
Qy 241 EANYNMLPPLDR 253  
Db 241 EANYNMLPPLGK 253

RESULT 17  
US-09-925-301-1006  
; Sequence 1006, Application US/09925301  
; Patent No. US20020052308A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies  
; FILE REFERENCE: PA106  
; CURRENT APPLICATION NUMBER: US/09/925,301  
; PRIOR FILING DATE: 2001-08-10  
; PRIOR APPLICATION NUMBER: PCR/US00/05882  
; PRIOR FILING DATE: 2000-03-08  
; PRIOR APPLICATION NUMBER: 60/124,270  
; PRIOR FILING DATE: 1999-03-12  
; NUMBER OF SEQ ID NOS: 1694  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO: 1006  
; LENGTH: 561  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-925-301-1006

Query Match 34.3%; Score 1125; DB 9; Length 561;  
Best Local Similarity 44.0%; Pred. No. 3,9e-100; Indels 38; Gaps 16;  
Matches 232; Conservative 87; Mismatches 170;

Qy 88 GNGEGGRP--YPMTEAERVDQAYRENGFNIVYSDKISLNRSLPDIRHPCNSKRYLETLP 145  
Db 59 GPGEMGRKVVIPKEDQKMKEMFKINQNLMASEMIALNRSLPDRLEGCKTKVYPNLP 118  
Qy 146 NTSIIIPFHNEGSSLLRTVHSLVNSRPPELVAEIVLVDPSREHLKKPLEDYM-ALFP 204  
Db 119 TTSVVIYFHNEMAWSTLRTVHSLVNSRPHMLEIVLVDASERDFLKRPLESYVKLKV 178  
Qy 205 SVRLRTKKEGRLRTMTLGSVATGVTIFLDSHCANYNMLPPLDLRIARNKTIYCP 264  
Db 179 PVHVRMEQSRGLIRALKGAASVKGQVITFLDAHCECTGWLLEPLARIKHDKRTVCP 238  
Qy 265 MIDVIDHDDPRYETQAG--DAMRGAFDMEYKRIPIRP---ELQKADSPDPSPVMAAG 320  
Db 239 IIVDISDTFEY--MAQSDMTYGGFNKLNFRMYVPQREMDRKGDTLPVRTPYMAAG 296  
Qy 321 LFAVDRKFMELGVDGLBIWGEQYEISFKVMCCGRMEDIPCSRVGHIYRKTYPVYK 380  
Db 297 LFSIDRDFQELIGTYDAGMDIWGENSELISPRIMQCGTLEIVTCSHVGHVFRKATPYTF 356  
Qy 381 PAGVS--LARNLKRAVAEVMDEYAEITYQRREYRHLISAGDVAVQKLRSSLNCKSFKWF 438  
Db 357 FGGTGQIINKNNRRILAEVWMDPEFTFYIISPGVTKVYDGDISRGLRHKLQCRPFSWY 416  
Qy 439 MTKIAMD--LPKFPYVPEPRPAAAMGEIRNVGTGLCAOTTKHGALGSLRLLEGCVRGGEAA 496  
Db 417 LENIYPSQIPIRHH-----FSLGEIRNVETNOCLDMMARKENKVGIFNC-HGMG-- 465  
Qy 497 WNNNQVFTTWRREDIRPDPOHTKKFCFPAISHTSPVTLVDCHSKMKNQMLKYRKDK-TL 555  
Db 466 --GNQVFSYTNKKEIRTD-----LCIDVSKLNGPVTMLKCHNLKGNQMLEYDVPVLT 517  
Qy 556 YHVPVSGCMD-CSESDHRI-FMNTCNPSLSLTQQLFPHNTSVLEKFP 600  
Db 518 QHVNNSOCLDKATEBDSQVPSIRDCN-GSRSQQLLR--NVTLPBIF 561

RESULT 18  
US-10-001-851-24  
; Sequence 24, Application US/10001851  
; Publication No. US20020115628A1  
; GENERAL INFORMATION:  
; APPLICANT: MEYERS, Rachel A.  
; APPLICANT: WILLIAMSON, Mark  
; TITLE OF INVENTION: Uses Thereof  
; FILE REFERENCE: 10147-56U1  
; CURRENT APPLICATION NUMBER: US/10/001,851  
; PRIOR FILING DATE: 2001-11-20  
; PRIOR APPLICATION NUMBER: US 60/249,939  
; PRIOR FILING DATE: 2000-11-20  
; NUMBER OF SEQ ID NOS: 29  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO: 24  
; LENGTH: 559  
; TYPE: PRT  
; ORGANISM: Sus sp.  
US-10-001-851-24

Query Match 34.1%; Score 1117; DB 13; Length 559;  
Best Local Similarity 43.6%; Pred. No. 2.4e-99;  
Matches 230; Conservative 90; Mismatches 169; Indels 38; Gaps 16;

Qy 88 GNGEGGRP--YPMTEAERVDQAYRENGFNIVYSDKISLNRSLPDIRHPCNSKRYLETLP 145  
Db 57 GPGEMGRKVVIPKEDQKMKEMFKINQNLMASEMIALNRSLPDRLEGCKTKVYPNLP 116  
Qy 146 NTSIIIPFHNEGSSLLRTVHSLVNSRPPELVAEIVLVDPSREHLKKPLEDYM-ALFP 204  
Db 117 TTSVVIYFHNEMAWSTLRTVHSLVNSRPHMLEIVLVDASERDFLKRPLESYVKLKV 176  
Qy 205 SVRLRTKKEGRLRTMTLGSVATGVTIFLDSHCANYNMLPPLDLRIARNKTIYCP 264  
Db 177 PVHVRMEQSRGLIRALKGAASVKGQVITFLDAHCECTGWLLEPLARIKHDKRTVCP 236  
Qy 265 MIDVIDHDDPRYETQAG--DAMRGAFDMEYKRIPIRP---ELQKADSPDPSPVMAAG 320  
Db 237 IIVDISDTFEY--MAQSDMTYGGFNKLNFRMYVPQREMDRKGDTLPVRTPYMAAG 294  
Qy 321 LFAVDRKFMELGVDGLBIWGEQYEISFKVMCCGRMEDIPCSRVGHIYRKTYPVYK 380  
Db 295 LFSIDRDFQELIGTYDAGMDIWGENSELISPRIMQCGTLEIVTCSHVGHVFRKATPYTF 354  
Qy 381 PAGVS--LARNLKRAVAEVMDEYAEITYQRREYRHLISAGDVAVQKLRSSLNCKSFKWF 438  
Db 355 FGGTGQIINKNNRRILAEVWMDPEFTFYIISPGVTKVYDGDISRGLRHKLQCRPFSWY 414  
Qy 439 MTKIAMD--LPKFPYVPEPRPAAAMGEIRNVGTGLCAOTTKHGALGSLRLLEGCVRGGEAA 496  
Db 415 LENIYPSQIPIRHH-----FSLGEIRNVETNOCLDMMARKENKVGIFNC-HGMG-- 463  
Qy 497 WNNNQVFTTWRREDIRPDPOHTKKFCFPAISHTSPVTLVDCHSKMKNQMLKYRKDK-TL 555  
Db 464 --GNQVFSYTNKKEIRTD-----LCIDVSKLNGPVTMLKCHNLKGNQMLEYDVPVLT 515  
Qy 556 YHVPVSGCMD-CSESDHRI-FMNTCNPSLSLTQQLFPHNTSVLEKFP 600  
Db 516 QHVNNSOCLDKATEBDSQVPSIRDCS-GSRSQQLLR--NVTLPBIF 559

RESULT 19  
US-10-001-851-21  
; Sequence 21, Application US/10001851  
; Publication No. US20020115628A1  
; GENERAL INFORMATION:  
; APPLICANT: MEYERS, Rachel A.  
; APPLICANT: WILLIAMSON, Mark  
; TITLE OF INVENTION: 47169 and 33935, No. US20020115628A1 Human Glycosyl Transferase



US-10-205-219-76

Query Match 34.0%; Score 1116; DB 14; Length 559;  
 Best Local Similarity 43.6%; Pred. No. 2.9e-99;  
 Matches 230; Conservative 88; Mismatches 171; Indels 38; Gaps 16;

QY 88 GNGEGRP--YPMTDARVDOAYRENGFNIVSDKISLNSRLPDIRHNCNSKRYLETLP 145  
 DB 57 GPGEMGRVVIIPKEDQEKMEKFKINQFNLMASEMIALNSRLPVRLEGCKTKYVPDNL 116  
 QY 146 NTSIIIPFHNMGSSLIARTVSHVNSRPPELVAEIVLVDDPSDRHKKPLEDYM-ALFP 204  
 DB 117 TTSVVIIFHNHAWSTLIRTVSHVNSRPRHMEIEIVLVDDASERDFPKRPLESVYKLV 176  
 QY 205 SVRLIRTKREGILIRTMIGASVATGVTITFLDSHCANVMWLPPLDIRIARNKTIYCP 264  
 DB 177 PVHVIIRMEQSGILIRALKGAASRGVITFLDAHCECTVGWLEPLARIKHDRITVCP 236  
 QY 265 MIDVIDHDDPRYETQAG-DAMRGAFDWMYYKRIPIPP--ELOKADSPDFESPVMAG 320  
 DB 237 IIDVISDTFEY--MAGSDMTYGGFNWKLNFWMYFVPOREMDRRKGRDRTLPVPTPMAG 294  
 QY 321 LFAVDRKFMELGIDPLGELIWGEGVIEISFKWMCGRMEDIPCSRVGHYRKVPYKV 380  
 DB 295 LFSIDRDFQEIIGTYDAGMDIWGEGNLEISFRIWQCGGTLEIVTCSHVGFRRATPYTF 354  
 QY 381 PAGVS--LAANLKVAEVMDEVAEYIYQRRPEVRLHAGDVAVQKLRSLNCKSKFMF 438  
 DB 355 PGGTGQIINKNNRLAEVWMEFKNFYIISPGVTKYDGDISRGVLRHLQCKPFSWY 414  
 QY 439 MTKIAMD--LPKFPYVPEPPAAWGEIRNVGTGLCADTKHAGLGSPLRLGCVVRGEAA 496  
 DB 415 LENIYPSQIPRHY-----FSLGEIRNVETNOCLDNMARKENKVGIFNC-HGMG--- 463  
 QY 497 WNNQVFTFTREDIRPGDPQHTKKCFPDASISHTSPVLYDCHSMKNQMLKTKRDK-TL 555  
 DB 464 --GNQVSYTANKSIRTD-----LCLDVSKLNGPVTMLKCHLKNQOLMEYDPVLT 515  
 QY 556 YHPVSGCMD-CSESDBRI-FMNTCNPSLSLTOQWLFHTNSTVLEKF 600  
 DB 516 QHVNNSOCLDKATBEDSQVPSIRDCT-GSRSQOMLNR--NVTLPRIIF 559

RESULT 22

US-10-001-851-22

Sequence 22, Application US/10001851

Publication No. US20020115628A1

GENERAL INFORMATION:

APPLICANT: MEYERS, Rachel A.

APPLICANT: WILLIAMSON, Mark

TITLE OF INVENTION: 47169 and 33935, No. US20020115628A1el Human Glycosyl Transferase

FILE REFERENCE: 10147-5601

CURRENT APPLICATION NUMBER: US/10/001,851

CURRENT FILING DATE: 2001-11-20

PRIOR APPLICATION NUMBER: US 60/249,939

NUMBER OF SEQ ID NOS: 29

SOFTWARE: PatentIn Ver. 2.1

SEQ ID NO 22

LENGTH: 559

TYPE: PRT

ORGANISM: Mus sp.

US-10-001-851-22

Query Match 34.0%; Score 1115; DB 13; Length 559;

Best Local Similarity 43.6%; Pred. No. 3.7e-99;

Matches 230; Conservative 87; Mismatches 172; Indels 38; Gaps 16;

QY 88 GNGEGRP--YPMTDARVDOAYRENGFNIVSDKISLNSRLPDIRHNCNSKRYLETLP 145  
 DB 57 GPGEMGRVVIIPKEDQEKMEKFKINQFNLMASEMIALNSRLPVRLEGCKTKYVPDNL 116

QY 146 NTSIIIPFHNMGSSLIARTVSHVNSRPPELVAEIVLVDDPSDRHKKPLEDYM-ALFP 204  
 DB 117 TTSVVIIFHNHAWSTLIRTVSHVNSRPRHMEIEIVLVDDASERDFPKRPLESVYKLV 176  
 QY 205 SVRLIRTKREGILIRTMIGASVATGVTITFLDSHCANVMWLPPLDIRIARNKTIYCP 264  
 DB 177 PVHVIIRMEQSGILIRALKGAASRGVITFLDAHCECTVGWLEPLARIKHDRITVCP 236  
 QY 265 MIDVIDHDDPRYETQAG-DAMRGAFDWMYYKRIPIPP--ELOKADSPDFESPVMAG 320  
 DB 237 IIDVISDTFEY--MAGSDMTYGGFNWKLNFWMYFVPOREMDRRKGRDRTLPVPTPMAG 294  
 QY 321 LFAVDRKFMELGIDPLGELIWGEGVIEISFKWMCGRMEDIPCSRVGHYRKVPYKV 380  
 DB 295 LFSIDRDFQEIIGTYDAGMDIWGEGNLEISFRIWQCGGTLEIVTCSHVGFRRATPYTF 354  
 QY 381 PAGVS--LAANLKVAEVMDEVAEYIYQRRPEVRLHAGDVAVQKLRSLNCKSKFMF 438  
 DB 355 PGGTGQIINKNNRLAEVWMEFKNFYIISPGVTKYDGDISRGVLRHLQCKPFSWY 414  
 QY 415 LENIYPSQIPRHY-----FSLGEIRNVETNOCLDNMARKENKVGIFNC-HGMG--- 463  
 QY 497 WNNQVFTFTREDIRPGDPQHTKKCFPDASISHTSPVLYDCHSMKNQMLKTKRDK-TL 555  
 DB 464 --GNQVSYTANKSIRTD-----LCLDVSKLNGPVTMLKCHLKNQOLMEYDPVLT 515  
 QY 556 YHPVSGCMD-CSESDBRI-FMNTCNPSLSLTOQWLFHTNSTVLEKF 600  
 DB 516 QHVNNSOCLDKATBEDSQVPSIRDCT-GSRSQOMLNR--NVTLPRIIF 559

RESULT 23

US-10-001-851-23

Sequence 23, Application US/10001851

Publication No. US20020115628A1

GENERAL INFORMATION:

APPLICANT: MEYERS, Rachel A.

APPLICANT: WILLIAMSON, Mark

TITLE OF INVENTION: 47169 and 33935, No. US20020115628A1el Human Glycosyl Transferase

FILE REFERENCE: 10147-5601

CURRENT APPLICATION NUMBER: US/10/001,851

CURRENT FILING DATE: 2001-11-20

PRIOR APPLICATION NUMBER: US 60/249,939

NUMBER OF SEQ ID NOS: 29

SOFTWARE: PatentIn Ver. 2.1

SEQ ID NO 23

LENGTH: 559

TYPE: PRT

ORGANISM: Bos sp.

US-10-001-851-23

Query Match 34.0%; Score 1115; DB 13; Length 559;

Best Local Similarity 43.6%; Pred. No. 3.7e-99;

Matches 230; Conservative 89; Mismatches 170; Indels 38; Gaps 16;

QY 88 GNGEGRP--YPMTDARVDOAYRENGFNIVSDKISLNSRLPDIRHNCNSKRYLETLP 145  
 DB 57 GPGEMGRVVIIPKEDQEKMEKFKINQFNLMASEMIALNSRLPVRLEGCKTKYVPDNL 116  
 QY 146 NTSIIIPFHNMGSSLIARTVSHVNSRPPELVAEIVLVDDPSDRHKKPLEDYM-ALFP 204  
 DB 117 TTSVVIIFHNHAWSTLIRTVSHVNSRPRHMEIEIVLVDDASERDFPKRPLESVYKLV 176  
 QY 205 SVRLIRTKREGILIRTMIGASVATGVTITFLDSHCANVMWLPPLDIRIARNKTIYCP 264  
 DB 177 PVHVIIRMEQSGILIRALKGAASRGVITFLDAHCECTVGWLEPLARIKHDRITVCP 236  
 QY 265 MIDVIDHDDPRYETQAG-DAMRGAFDWMYYKRIPIPP--ELOKADSPDFESPVMAG 320

Db 237 IIDVSDTTEY--MAGSDMTYGGFNWKLNRWYVPVQREMDRRKGRDRTLPVPTPTMAGG 294  
Qy 321 LFAVDRKMFELGSDYDGLIHWGEOYEISFKVMCGGEMEDIDCSRGVHYYRKYVYK 380  
Db 295 LFSIDRNYFEIGTYDAGMDIWGEMLEISFRIMQCGGTLEIVTCSHVGHVFRATPYTF 354  
Qy 381 PAGVS--LAARLKKVAEVMDEYAEYIYQRRPEYRHLISAGDVAOVOKLBSLACKSKPWF 438  
Db 355 PGGTGOIINKNNRRLAEVWMDYFKNFYIISPGVTKVDYDISRLRLKHLQCRPSWY 414  
Qy 439 MTKIAMD--LPKFYPPVBPAAAMGEIRNVTGTCADTKHAGLSPRLBSCVGRGEAA 496  
Db 415 LENIYPSQIPRHY-----FSLGRIINVTETNOCLDMARKENKYGIFNC-HGMG--- 463  
Qy 497 WNNQVFTFTWRREDIRPGDPQHTKKFCFDAISHTSPVTLVDCHSMKNQMLKRYKDK-TL 555  
Db 464 --GNQVFSYANKKEIRTD-----LCIDVSKLNGPVTMLKCHLKNQMLWEYDPVXLTL 515  
Qy 556 YHPVSGCMD-CSESDHRI-FMNTCNPSLSLQOMLFHTNSTVLEKF 600  
Db 516 OHVNSNOCLDKATEDEQVPSIRDCS-GSRSQOMLR--NVTLPETIF 559

RESULT 24  
US-10-001-851-20

Sequence 20, Application US/10001851  
Publication No. US20020115628A1  
GENERAL INFORMATION:  
APPLICANT: MEYERS, Rachel A.  
APPLICANT: WILLIAMSON, Mark  
TITLE OF INVENTION: 47169 and 33935, No. US20020115628A1 Human Glycosyl Transferase  
TITLE OF INVENTION: Uses thereof  
FILE REFERENCE: 10147-5601  
CURRENT APPLICATION NUMBER: US/10/001, 851  
CURRENT FILING DATE: 2001-11-20  
PRIORITY APPLICATION NUMBER: US 60/249, 939  
PRIORITY FILING DATE: 2000-11-20  
NUMBER OF SEQ ID NOS: 29  
SOFTWARE: Patent In Ver. 2.1  
SEQ ID NO 20  
LENGTH: 559  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-001-851-20

Query Match 34.0%; Score 113; DB 13; Length 559;  
Best Local Similarity 43.6%; Pred. No. 5,8e-99;  
Matches 230; Conservative 87; Mismatches 172; Indels 38; Gaps 16;

Qy 88 GNGEGRP--YPMTDARVDOAVRENGFNIVSDKISLNSLPDIRHPNCKSKRYLETLP 145  
Db 57 GPGEMGKAVLIIPKDDQKMEKELKINQFNLMASDILNLSLPVRLGCKTKYPPDLPL 116  
Qy 146 NTSIIIPFHNWSSLLRTVSHVNLNRSPPRLVAEIVLVDDPSDREHLKKPLEDYMA-LFP 204  
Db 117 TTSVIAIFHNWASTLRTVSHVNLNRSPPRLVAEIVLVDDPSDREHLKKPLEDYMA-LFP 176  
Qy 205 SVRLTKKEGGLIRTMGLASVATGVTITFLDSHCANVWMLPPLDIRIARNRKTYVCP 264  
Db 177 PVAHYIRMEGSGILRLKGAASVKGQVITFLDAHCECTGVMLPLARIHMDRITVCP 236  
Qy 265 MIDVIHDDRYETQAG-DAMRGAFDWMYKRIPIPP--ELQKADSPDPFSPVMAAG 320  
Db 237 IIDVSDTTEY--MAGSDMTYGGFNWKLNRWYVPVQREMDRRKGRDRTLPVPTPTMAGG 294  
Qy 321 LFAVDRKMFELGSDYDGLIHWGEOYEISFKVMCGGEMEDIDCSRGVHYYRKYVYK 380  
Db 295 LFSIDRNYFEIGTYDAGMDIWGEMLEISFRIMQCGGTLEIVTCSHVGHVFRATPYTF 354  
Qy 381 PAGVS--LAARLKKVAEVMDEYAEYIYQRRPEYRHLISAGDVAOVOKLBSLACKSKPWF 438  
Db 355 PGGTGOIINKNNRRLAEVWMDYFKNFYIISPGVTKVDYDISRLRLKHLQCRPSWY 414

Qy 439 MTKIAMD--LPKFYPPVBPAAAMGEIRNVTGTCADTKHAGLSPRLBSCVGRGEAA 496  
Db 415 LENIYPSQIPRHY-----FSLGRIINVTETNOCLDMARKENKYGIFNC-HGMG--- 463  
Qy 497 WNNQVFTFTWRREDIRPGDPQHTKKFCFDAISHTSPVTLVDCHSMKNQMLKRYKDK-TL 555  
Db 464 --GNQVFSYANKKEIRTD-----LCIDVSKLNGPVTMLKCHLKNQMLWEYDPVXLTL 515  
Qy 556 YHPVSGCMD-CSESDHRI-FMNTCNPSLSLQOMLFHTNSTVLEKF 600  
Db 516 OHVNSNOCLDKATEDEQVPSIRDCN-GSRSQOMLR--NVTLPETIF 559

RESULT 25  
US-10-292-896-62

Sequence 62, Application US/10292896  
Publication No. US20030186850A1  
GENERAL INFORMATION:  
APPLICANT: HASSAN, Helle  
APPLICANT: REIS, Celso A.  
APPLICANT: BENNETT, Eric P.  
TITLE OF INVENTION: METHODS TO IDENTIFY AGENTS MODULATING FUNCTIONS OF POLYPEPTIDE GAV  
TITLE OF INVENTION: TRANSFERASES, PHARMACEUTICAL COMPOSITIONS COMPRISING SUCH AGENTS  
TITLE OF INVENTION: SUCH AGENTS FOR PREPARING MEDICATIONS  
FILE REFERENCE: 4305/1H154-US3  
CURRENT APPLICATION NUMBER: US/10/292, 896  
CURRENT FILING DATE: 2003-03-17  
PRIORITY APPLICATION NUMBER: US 60/425, 204  
PRIORITY FILING DATE: 2002-11-08  
PRIORITY APPLICATION NUMBER: PCT/DK01/00328  
PRIORITY FILING DATE: 2001-05-10  
PRIORITY APPLICATION NUMBER: US 60/203, 331  
PRIORITY FILING DATE: 2000-05-11  
NUMBER OF SEQ ID NOS: 130  
SOFTWARE: Patent In version 3.1  
SEQ ID NO 62  
LENGTH: 556  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-292-896-62

Query Match 33.7%; Score 1106; DB 14; Length 556;  
Best Local Similarity 43.8%; Pred. No. 2,8e-98;  
Matches 227; Conservative 90; Mismatches 165; Indels 36; Gaps 15;

Qy 88 GNGEGRP--YPMTDARVDOAVRENGFNIVSDKISLNSLPDIRHPNCKSKRYLETLP 145  
Db 56 GPGEMGKAVLIIPKDDQKMEKELKINQFNLMASDILNLSLPVRLGCKTKYPPDLPL 115  
Qy 146 NTSIIIPFHNWSSLLRTVSHVNLNRSPPRLVAEIVLVDDPSDREHLKKPLEDYMA-LFP 204  
Db 116 NTSVIAIFHNWASTLRTVSHVNLNRSPPRLVAEIVLVDDPSDREHLKKPLEDYMA-LFP 175  
Qy 205 SVRLTKKEGGLIRTMGLASVATGVTITFLDSHCANVWMLPPLDIRIARNRKTYVCP 264  
Db 176 PVAHYIRMEGSGILRLKGAASVKGQVITFLDAHCECTGVMLPLARIHMDRITVCP 235  
Qy 265 MIDVIHDDRYETQAG-DAMRGAFDWMYKRIPIPP--ELQKADSPDPFSPVMAAG 320  
Db 236 IIDVSDTTEY--MAGSDMTYGGFNWKLNRWYVPVQREMDRRKGRDRTLPVPTPTMAGG 293  
Qy 321 LFAVDRKMFELGSDYDGLIHWGEOYEISFKVMCGGEMEDIDCSRGVHYYRKYVYK 380  
Db 294 LFSIDRNYFEIGTYDAGMDIWGEMLEISFRIMQCGGTLEIVTCSHVGHVFRATPYTF 353  
Qy 381 PAGVS--LAARLKKVAEVMDEYAEYIYQRRPEYRHLISAGDVAOVOKLBSLACKSKPWF 438  
Db 354 PGGTGOIINKNNRRLAEVWMDYFKNFYIISPGVTKVDYDISRLRLKHLQCRPSWY 413  
Qy 439 MTKIAMDLPKFYPPVBPAAAM--GEIRNVTGTCADTKHAGLSPRLBSCVGRGEAA 496  
Db 414 LENI-----YDPSQIPRHYISLGEIRNVTETNOCLDMARKENKYGIFNC-HGMG--- 462

OY 497 WNNMVFETFTWRBDRPGDPQHTKKCFDAISHSPVTLTDCHSKGNQWRYRDK-TL 555  
 Db 463 --GNQVPSYTDKERTD-----LCUDVSRINGPVIMLKCHMRGNQWMEYDAERLTL 514  
 OY 556 YHPVSGSCMD-CSES DHRI-FMNTCNPSSLTQOWLFEEH 591  
 Db 515 RHVNSNOCLDEPSEEDKVPMTQDCS-GSRSQOWILLRN 551

Search completed: November 22, 2004, 13:42:20  
 Job time : 58 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

## OM protein - protein search, using sw model

Run on: November 22, 2004, 13:38:52 ; Search time 21 Seconds

(Without alignments)  
2762.796 Million cell updates/sec

Title: US-10-001-851-2

Perfect score: 3278  
Sequence: 1 MRKKERRLQVALVALVALV.....TQGMLEHTNISTYLEKRN 603

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database :

1: PIR 79: \*  
2: PIR2: \*  
3: PIR3: \*  
4: PIR4: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1392.5	42.5	684	2	T26930	hypothetical prote
2	1392	42.5	276	2	T12552	hypothetical prote
3	1125	34.3	559	2	JC4223	polypeptide N-acet
4	1115	34.0	559	2	A45987	polypeptide N-acet
5	1113	34.0	601	2	T42251	polypeptide N-acet
6	1083	33.0	624	2	T42247	polypeptide N-acet
7	1079	32.9	626	2	T42246	polypeptide N-acet
8	1078.5	32.9	623	2	T42245	probable polypepti
9	988.5	30.2	563	2	A88515	polypeptide N-acet
10	988.5	30.2	612	2	T42243	probable polypepti
11	968	29.5	571	2	T42249	polypeptide N-acet
12	949	29.0	571	2	T37405	polypeptide N-acet
13	948.5	28.9	618	2	T42248	polypeptide N-acet
14	938	28.6	562	2	T42250	polypeptide N-acet
15	934	28.5	633	2	JC5247	polypeptide N-acet
16	906	27.6	579	2	T31549	polypeptide N-acet
17	883	26.9	589	2	T42244	probable polypepti
18	801	24.4	605	2	T27397	hypothetical prote
19	632	19.3	421	2	T42252	polypeptide N-acet
20	147	4.5	1044	2	H97186	glycosyltransferas
21	139	4.2	306	2	D87531	glycosyl transfera
22	138.5	4.2	308	2	H87306	glycosyl transfera
23	135	4.1	328	2	B84253	electroglycan bios
24	135.5	4.1	312	2	S74669	hypothetical prote
25	135.5	4.1	318	2	D87506	glycosyl transfera
26	130	4.0	322	2	AC2023	hypothetical prote
27	130	4.0	787	2	H98163	hypothetical prote
28	128	3.9	470	2	C70641	hypothetical prote
29	126	3.8	313	2	AI2404	hypothetical prote

30	124	3.8	362	2	G75191	dolichol-phosphate
31	124	3.8	972	2	T09595	glucuronosyltransf
32	123	3.8	334	1	G71153	hypothetical prote
33	122	3.7	316	2	AB2868	UDP-hexose transfe
34	122	3.7	316	2	F97644	probable glycosyl
35	120	3.7	260	2	E90984	glycosyl transfera
36	120	3.7	260	2	H85829	glycosyl transfera
37	119.5	3.6	416	2	AE1499	conserved hypotnet
38	118.5	3.6	288	2	B75096	glycosyl transfera
39	118.5	3.6	416	2	AB1141	probable glucosami
40	117.5	3.6	245	2	D87307	glycosyl transfera
41	117	3.6	343	2	AI2091	glycosyltransferas
42	116	3.5	251	2	AI2106	hypothetical prote
43	115.5	3.5	269	2	AI3123	glycosyltransferas
44	115.5	3.5	392	2	H69814	hypothetical prote
45	115.5	3.5	477	1	US0589	endo-1,4-beta-xyla
46	114	3.5	333	2	AH2026	hypothetical prote
47	113.5	3.5	301	2	F95205	glycosyl transfera
48	111.5	3.4	211	2	A75176	dolichyl-phosphate
49	111.5	3.4	283	2	T00099	glycosyltransferas
50	111.5	3.4	383	2	T34603	xylinase A - Strept
51	111	3.4	2774	2	A43359	microtubule-associ
52	110	3.4	215	1	D71038	hypothetical prote
53	108	3.3	233	2	H97103	probable glycosylt
54	108	3.3	274	2	B81319	probable glycosylt
55	108	3.3	353	2	H71223	probable dolichol-
56	108	3.3	448	2	T01814	hypothetical prote
57	107.5	3.3	332	2	T44647	glycosyl transfera
58	107.5	3.3	371	2	H69290	dolichol-P-glucose
59	105.5	3.2	319	2	A70715	hypothetical prote
60	105	3.2	294	2	D64175	glycosyltransferas
61	104.5	3.2	309	2	B87550	glycosyl transfera
62	104.5	3.2	581	2	B69322	dolichol-P-glucose
63	104.5	3.2	615	2	C97723	aspartate-tRNA lig
64	104.5	3.2	745	1	A70458	phosphoribosylform
65	104	3.2	277	2	C97626	dolichol-phosphate
66	104	3.2	279	2	AE2849	probable glycosylt
67	103.5	3.2	344	2	AC0974	probable glycosylt
68	103.5	3.2	348	2	D95977	glucosyltransferas
69	103.5	3.2	348	2	S39958	exoO protein - Rhl
70	103.5	3.2	420	2	D69769	cellulose synthase
71	103.5	3.2	570	2	S62627	agglutinin I prepu
72	102.5	3.1	564	1	PLCSAG	agglutinin precurs
73	102	3.1	291	2	F95015	glycosyl transfera
74	102	3.1	295	2	AE2367	hypothetical prote
75	102	3.1	317	2	H97888	glycosyl transfera
76	102	3.1	328	2	F95158	glycosyl transfera
77	102	3.1	459	2	G69075	nitrogenase molybd
78	101.5	3.1	331	2	G75099	sugar transferase
79	101.5	3.1	344	2	A70037	capsular polysacch
80	101.5	3.1	335	2	A48755	hyaluronan synthas
81	101.5	3.1	419	2	A53100	hyaluronate synth
82	101	3.1	301	2	AI1920	hypothetical prote
83	101	3.1	328	2	F98024	raffinose-raffinof
84	101	3.1	380	2	S76171	hypothetical prote
85	101	3.1	637	2	A29609	reticacycline resis
86	101	3.1	1143	2	S46122	SNF2 protein homol
87	100.5	3.1	318	2	AC2152	hypothetical prote
88	100.5	3.1	412	2	AC2152	dolichyl-phosphate
89	100.5	3.1	724	2	T04340	beta-galactosidase
90	100.5	3.1	1012	2	B90389	conserved hypotnet
91	100	3.1	307	2	AH1366	fructose-1-phosph
92	100	3.1	313	2	AI2202	hypothetical prote
93	99.5	3.0	303	2	F81318	probable galactosy
94	99.5	3.0	318	2	AE5043	probable pre-mRNA
95	99	3.0	318	2	AG2189	hypothetical prote
96	99	3.0	327	2	G89990	hypothetical prote
97	99	3.0	475	2	T35697	arabinoturanosidas
98	99	3.0	664	2	T47481	receptor-like prot
99	98.5	3.0	739	2	T12983	hypothetical prote
100	98	3.0	257	2	E84107	teichuronic acid b

## ALIGNMENTS

## RESULT 1

T26930  
 hypothetical protein Y45F10D.3 - *Caenorhabditis elegans*  
 C/Species: *Caenorhabditis elegans*  
 C/Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004  
 C/Accession: T26930  
 R/Murray A.  
 submitted to the EMBL Data Library, January 1998  
 A/Reference number: Z2028  
 A/Accession: T26930  
 A/Status: preliminary; translated from GB/EMBL/DBJ  
 A/Molecule type: DNA  
 A/Residues: 1-684 <MTL>  
 A/Cross-references: UNIPROT:O45947, EMBL:AL021492, PIDN:CAA16378.1, GSPDB:GN00022, CESP:  
 A/Experimental source: clone Y45F10D  
 C/Genetics:  
 A/Gene: CESP:Y45F10D.3  
 A/Map position: 4  
 A/Intons: 61/3; 112/1; 142/3; 178/3; 230/3; 289/2; 611/2  
 C/Superfamily: polypeptide N-acetylglucosaminyltransferase

Query Match 42.5%; Score 1392.5; DB 2; Length 684;  
 Best Local Similarity 48.8%; Pred. No. 1.6e-104;  
 Matches 279; Conservative 87; Mismatches 175; Indels 31; Gaps 14;

```

QY 49 PAAGGSHRQKTPFLDQGLKDMHDKAIRDPAQVNGEGCRPTPTMTDAERVD-- 106
DB 121 PPALGGEALDPFEKYRGH-EKIK-WEDEAAYEKREKREGGEWCKPVKLPEDKEVEEA 178
QY 107 --AYRENGFNIVSDKISLNSLPDIRHPNCNKKYLETLPNTSIIIFHNHGSMLLT 164
DB 179 LSLYKANGYAYIYSDMISLNSIKDIRKECKNMYSKALPTVSYTFPHEHNSTLIRS 238
QY 165 VHSVLSNSPPELVVAIYLVDFDSDEHLKPLBDM--ALFPVYRLITKKRGLIRTR 221
DB 239 VYSVINSPPLELKEIILVDFSEKPAALRQLEDFLKNKIKDHIKVLRTKKRGLINGR 298
QY 222 MLGASVATGVTITPLDSHCENANWMLPPLDLRIANRKTIVCPMLDIVDHDHDFRYETQAG 281
DB 299 QLGADATGELIILFLDAISEANYWMLPPLDLPIEDYRTVCPFDVIDCEITYEVRP-D 357
QY 282 DAMGAFDMENYKRIPIPELOKADPDPFSPPMAGLFAVDKRMFELGGYDPGEI 341
DB 358 EGANGSEPMANRYKRLPIITK-DRESPTKPRNSPMAGGYFAISKMFELGGIDEGLDI 416
QY 342 WGEQYEISFYVMCGGEMEDIPCSRGVHIYR-KYVPYK-VPAGVSLANLKRVAEVMMD 399
DB 417 WGEQYEISFYVMCGGEMEDIPCSRGVHIYR-KYVPYK-VPAGVSLANLKRVAEVMMD 476
QY 400 EYAAIYIQRRPEYHLSGDVAVOGKLASSLNCSPFKFMTKILAMDLPKFPVPEPPAAA 459
DB 477 DYKETLVKRRGVGADADLKMKGIREKLQCKSPDFMFEIADFDDKYPVAEPPKSA 536
QY 460 WGEIRNVTGTCADPTKHALGSPLEEGCVR-----GRGEAAMNNQVTFPMREDIRPGD 515
DB 537 EGEIRNVTGTCADPTKHALGSPLEEGCVR-----QDLRLRMHDIR-- 588
QY 516 POKTKKFCFDAISHT--SPVTLVYDCHSMKNQMLWYR-KDXTLYHPVSGSCMDSESDDR 572
DB 589 --KGRKICFDCSTSVDAKAVILFDCHSMKNQMLFYKVAQVQIYHPISGQCLTADENKSG 646
QY 573 -IFMNTCNPSLTOQMLEFHNTSTVLEKFNRR 603
DB 647 FLHMKKCDSSSLQKMAWQTVDNELLETRQAN 678

```

## RESULT 2

T12552  
 hypothetical protein DKFZp586H0623.1 - human (fragments)  
 C/Species: *Homo sapiens* (man)

C/Date: 23-Jul-1999 #sequence\_revision 23-Jul-1999 #text\_change 09-Jul-2004  
 C/Accession: T12552  
 R/Amstrong, W.; Winkner, U.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.  
 submitted to the Protein Sequence Database, June 1999  
 A/Reference number: Z17527  
 A/Accession: T12552  
 A/Status: preliminary  
 A/Molecule type: mRNA  
 A/Residues: 1-150,151-276 <ANS>  
 A/Cross-references: UNIPROT:Q9Y4M4, EMBL:AL096739  
 A/Experimental source: adult uterus, clone DKFZp586H0623  
 A/Note: the cDNA sequence contains a -1 frameshift near codon 150  
 C/Genetics:  
 A/Note: DKFZp586H0623.1  
 C/Superfamily: polypeptide N-acetylglucosaminyltransferase

Query Match 42.5%; Score 1392; DB 2; Length 276;  
 Best Local Similarity 97.3%; Pred. No. 4.6e-105;  
 Matches 250; Conservative 3; Mismatches 2; Indels 2; Gaps 1;

```

QY 349 ISFKVMCGGEMEDIPCSRGVHIYRKYVPYKVPAGVSLA--RNLKRYAEVMDYAEYIY 406
DB 20 LAMQVMCGGEMEDIPCSRGVHIYRKYVPYKVPAGVSLARVTLKRYAEVMDYAEYIY 79
QY 407 QRRPEYHLSAGDVAVOGKLASSLNCSPFKFMTKILAMDLPKFPVPEPPAAAAGEIRNV 466
DB 80 QRRPEYHLSAGDVAVOGKLASSLNCSPFKFMTKILAMDLPKFPVPEPPAAAAGEIRNV 139
QY 467 GTGICATTKIGALGSPRLRECVRGREAAAMNNQVTFPMREDIRGDDPQHTKKFCEDA 526
DB 140 GTGICATTKIGALGSPRLRECVRGREAAAMNNQVTFPMREDIRGDDPQHTKKFCEDA 199
QY 527 ISHSPVTLVYDCHSMKNQMLWYR-KDXTLYHPVSGSCMDSESDDRIFMNTCNPSLTOQ 586
DB 200 ISHSPVTLVYDCHSMKNQMLWYR-KDXTLYHPVSGSCMDSESDDRIFMNTCNPSLTOQ 259
QY 587 WLFPHNTSTVLEKFNRR 603
DB 260 WLFPHNTSTVLEKFNRR 276

```

## RESULT 3

JC4223  
 polypeptide N-acetylglucosaminyltransferase (EC 2.4.1.41) - human  
 M/Alternate names: GalNAc-transferase  
 C/Species: *Homo sapiens* (man)  
 C/Date: 14-Nov-1995 #sequence\_revision 08-Feb-1996 #text\_change 09-Jul-2004  
 C/Accession: JC4223; I37404  
 R/Meurer, J.A.; Naylor, J.M.; Baker, C.A.; Thomsen, D.R.; Homa, F.L.; Elhammer, A.P.  
 J. Biochem. 118, 568-574, 1995  
 A/Title: cDNA cloning, expression, and chromosomal localization of a human UDP-GalNAc: p  
 A/Reference number: JC4223; MUID:96115928; PMID:8690729  
 A/Accession: JC4223  
 A/Molecule type: mRNA  
 A/Residues: 1-559 <MEU>  
 A/Cross-references: UNIPROT:Q10472  
 A/Experimental source: salivary gland  
 A/Note: The authors translated the codon AAT for residue 264 as Asp  
 R/White, T.; Bennett, E.P.; Paul, E.; Takio, K.; Stensen, T.; Bonding, N.; Clausen, H.  
 J. Biol. Chem. 270, 24156-24165, 1995  
 A/Title: Purification and cDNA cloning of a human UDP-GalNAc:polypeptide N-Acetylglucto-  
 A/Reference number: I37404; MUID:96025800; PMID:7592619  
 A/Accession: I37404  
 A/Status: preliminary; translated from GB/EMBL/DBJ  
 A/Molecule type: mRNA  
 A/Residues: 1-559 <RES>  
 A/Cross-references: EMBL:X85018, NID:9971458, PIDN:CAA59380.1, PID:9971459  
 C/Comment: This enzyme catalyzes the initial reaction in O-linked (mucin type) oligosacchi-  
 retory pathway.

C/Genetics:  
 A/Map position: 18  
 C/Superfamily: polypeptide N-acetylglucosaminyltransferase  
 C/Keywords: chromosomal protein; glycoprotein; glycosyltransferase; hexosyltransferase;

F:9-28/Domain: transmembrane #status predicted <TM>

F:29-559/Domain: endoplasmic reticulum lumenal #status predicted <LM>

F:95,141,541,552/Binding site: carbohydrate (Asn) (covalent) #status predicted  
F:117,118,288/Binding site: carbohydrate (Thr) (covalent) #status predicted  
F:119/Binding site: carbohydrate (Ser) (covalent) #status predicted

Query Match 34.3%; Score 1125; DB 2; Length 559;

Best Local Similarity 44.0%; Pred. No. 5,66-83;  
Matches 232; Conservative 87; Mismatches 170; Indels 38; Gaps 16;

```

Qy 88 GNGEQGRP--YPMTDARVQAVRENGFNIVSDKISLNSLPDIRHPCNSKRYLETLP 145
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 57 GPGEMKRPVYIPKEDQKMKEMFKINQFNLMASMLNSLPVRLKRGCTKYYPNLP 116
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 146 NTSIIIPFNHNGWSSLRTVSHVNLNSRPPELVAVIVLVDPSDREHLKRPLEDYM-ALFP 204
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 117 TTSVVIIFHNHNAVSTLRTVSHVNLNSRPPELVAVIVLVDASERDPLKRPLESYVKGLKV 176
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 205 SVRLRTKREGLIRTMGLASVATGVTITFLSHCANVNWLPPLDRIARNKTIYCP 264
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 177 PVHVRMQRSGLRRLKGAASVKGQVITFLDHACECTGWLPEPLARIGHDKRTVCP 236
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 265 MIDVIDHDDPRYETQAG--DAMRGAFDWEMTKRPIPP---ELQKADPSDFESPVAAG 320
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 237 IIDVISDTEFY--MAGSDMTYGGFNKKLNFRWYVQREMRRKGRITLVPRTPTMAGG 294
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 321 LFAVDRKFMELGSDYDGLBIWGEQYEIFKVMCGRMEDIPCSRVGHIYRKTYPVYK 380
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 295 LFSIDRDFQEIIGYDAGMGIWGENLEISFRIMQCGGLEIYVCSHVGHVFRKATPTYF 354
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 381 PAGVS--LAENLRVAEVMDEVAEYIYQRRPEYRHLISAGDVAVQKTLRSSLNCKSPKWF 438
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 355 PGGTGQIINKNNRRLAEVWMDDEFKNFFIISPGVTKVDYGDISSRGLRHLQCRPSWY 414
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 439 MTKIAMD--LPKFPYVPPAPAAWGEIRNYGTGLCADTKGALGSLRLKGVNREBAA 496
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 415 LENIYPSQIPRHH-----FSLGEIRNVEYTNQCLDNMARKENKXGIFNC-HGMG--- 463
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 497 WNNMQVFTFWREDIRPGDPQHTKKCFDAISHTSPVTLVDCSHKMNQMLKRYKDK-TL 555
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 464 --GNQVFSYANKKEIRTD-----LCIDVSKLNGPVTMLKCHHLKQNLMEYDPAVLT 515
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 556 YHPVSGSCMD-CSESDHRI-FMNTCNPSLSLQQLFPHNTSVYLEKF 600
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 516 QHVNNSQCLDKATVEDSQVPSIRDCN-GSRSQQLLR--NVTLPETIF 559
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

#### RESULT 4

A45987 polypeptide N-acetylglactosaminyltransferase (EC 2.4.1.41) - bovine

C/Species: Bos primigenius taurus (cattle)  
C/Date: 03-May-1994 #sequence\_revision 03-May-1994 #text\_change 09-Jul-2004  
C/Accession: A45987; A48530

R/Hagen, F.L.; Hollander, T.; Lehman, D.J.; Thomsen, D.R.; Elhammer, A.P.  
J. Biol. Chem. 268, 12609-12616, 1993

A/Title: Isolation and expression of a cDNA clone encoding a bovine UDP-GalNAc:polypeptide  
A/Reference number: A45987; MUID:9328609; PMID:7685345

A/Accession: A45987

A/Status: preliminary

A/Molecule type: mRNA

A/Residues: 1-559 <HOM>

A/Residues: 1-559 <HOM>

A/Cross-references: UNIPROT:Q07537; GB:L07780; NID:g289411; PIDN:AAA30532.1; PID:g289412

A/Note: authors translated the codon GAT for residue 310 as Asn

R/Hagen, F.K.; VanWuyckhuysen, B.; Tabak, L.A.

J. Biol. Chem. 268, 18960-18965, 1993

A/Title: Purification, cloning, and expression of a bovine UDP-GalNAc: polypeptide N-ac

A/Reference number: A48530; MUID:93366815; PMID:8360184

A/Accession: A48530

A/Status: preliminary

A/Molecule type: mRNA

A/Residues: 41-559 <HAG>

A/Cross-references: GB:L17437; NID:g304258; PIDN:AAA68489.1; PID:g304259

C/Superfamily: polypeptide N-acetylglactosaminyltransferase

C/Keywords: colostrum; glycosyltransferase; hexosyltransferase

Query Match 34.0%; Score 1115; DB 2; Length 559;

Best Local Similarity 43.6%; Pred. No. 3,66-82;  
Matches 230; Conservative 89; Mismatches 170; Indels 38; Gaps 16;

```

Qy 88 GNGEQGRP--YPMTDARVQAVRENGFNIVSDKISLNSLPDIRHPCNSKRYLETLP 145
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 57 GPGEMKRPVYIPKEDQKMKEMFKINQFNLMASMLNSLPVRLKRGCTKYYPNLP 116
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 146 NTSIIIPFNHNGWSSLRTVSHVNLNSRPPELVAVIVLVDPSDREHLKRPLEDYM-ALFP 204
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 117 TTSVVIIFHNHNAVSTLRTVSHVNLNSRPPELVAVIVLVDASERDPLKRPLESYVKGLKV 176
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 205 SVRLRTKREGLIRTMGLASVATGVTITFLSHCANVNWLPPLDRIARNKTIYCP 264
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 177 PVHVRMQRSGLRRLKGAASVKGQVITFLDHACECTGWLPEPLARIGHDKRTVCP 236
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 265 MIDVIDHDDPRYETQAG--DAMRGAFDWEMTKRPIPP---ELQKADPSDFESPVAAG 320
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 237 IIDVISDTEFY--MAGSDMTYGGFNKKLNFRWYVQREMRRKGRITLVPRTPTMAGG 294
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 321 LFAVDRKFMELGSDYDGLBIWGEQYEIFKVMCGRMEDIPCSRVGHIYRKTYPVYK 380
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 295 LFSIDRDFQEIIGYDAGMGIWGENLEISFRIMQCGGLEIYVCSHVGHVFRKATPTYF 354
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 381 PAGVS--LAENLRVAEVMDEVAEYIYQRRPEYRHLISAGDVAVQKTLRSSLNCKSPKWF 438
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 355 PGGTGQIINKNNRRLAEVWMDDEFKNFFIISPGVTKVDYGDISSRGLRHLQCRPSWY 414
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 439 MTKIAMD--LPKFPYVPPAPAAWGEIRNYGTGLCADTKGALGSLRLKGVNREBAA 496
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 415 LENIYPSQIPRHH-----FSLGEIRNVEYTNQCLDNMARKENKXGIFNC-HGMG--- 463
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 497 WNNMQVFTFWREDIRPGDPQHTKKCFDAISHTSPVTLVDCSHKMNQMLKRYKDK-TL 555
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 464 --GNQVFSYANKKEIRTD-----LCIDVSKLNGPVTMLKCHHLKQNLMEYDPAVLT 515
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 556 YHPVSGSCMD-CSESDHRI-FMNTCNPSLSLQQLFPHNTSVYLEKF 600
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 516 QHVNNSQCLDKATVEDSQVPSIRDCS-GSRSQQLLR--NVTLPETIF 559
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

#### RESULT 5

T42251 polypeptide N-acetylglactosaminyltransferase (EC 2.4.1.41) 7 - Caenorhabditis elegans

C/Species: Caenorhabditis elegans  
C/Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 09-Jul-2004  
C/Accession: T42251

R/Hagen, F.K.; Nehrke, K.  
J. Biol. Chem. 273, 8268-8277, 1998

A/Title: cDNA cloning and expression of a family of UDP-N-acetyl-Dgalactosamine:Polypept

A/Reference number: 222126; MUID:98192620; PMID:9555933

A/Accession: T42251

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Residues: 1-601 <HAG>

A/Residues: 1-601 <HAG>

A/Cross-references: UNIPROT:O61397; EMBL:AF031841; NID:g3047202; PIDN:AA13677.1; PID:g3

A/Status: preliminary

A/Molecule type: mRNA

A/Residues: 1-601 <HAG>

A/Cross-references: UNIPROT:O61397; EMBL:AF031841; NID:g3047202; PIDN:AA13677.1; PID:g3

A/Status: preliminary

A/Molecule type: mRNA

A/Residues: 1-601 <HAG>

A/Cross-references: UNIPROT:O61397; EMBL:AF031841; NID:g3047202; PIDN:AA13677.1; PID:g3

A/Status: preliminary

A/Molecule type: mRNA

A/Residues: 1-601 <HAG>

A/Cross-references: UNIPROT:O61397; EMBL:AF031841; NID:g3047202; PIDN:AA13677.1; PID:g3

A/Status: preliminary

A/Molecule type: mRNA

A/Residues: 1-601 <HAG>

A/Cross-references: UNIPROT:O61397; EMBL:AF031841; NID:g3047202; PIDN:AA13677.1; PID:g3

A/Status: preliminary

A/Molecule type: mRNA

A/Residues: 1-601 <HAG>

A/Cross-references: UNIPROT:O61397; EMBL:AF031841; NID:g3047202; PIDN:AA13677.1; PID:g3

A/Status: preliminary

A/Molecule type: mRNA

Db 136 RTIPDIRBECKHWDYEKLPTVSVVVVFHNEGWTPLLRTHSVLSRPELLEQVWMD 195  
Qy 185 DFDREHLKKPLIEDYMLF-PSVRLRTKKEGHLIRTMGLASAVATGVTFLDSHCEN 243  
Db 196 DSDKPLHKEKLDKYTRFNKGKIVVTRQEGELINRSISAKSISGEVFLDAHGBVN 255  
Qy 244 VNNLPPLLDRIARNRKTIVCEMIDVIDHDPRYETQAGDA---MRGAFDWMYKRIPIP 300  
Db 256 TNNLPPLLIKRRKRWTVTVIDIGIDNSWEYSVSGPNAHSGIFEMQLYKEQIT 315  
Qy 301 PE--LQCADSDPESPESPMAGGLFAVDRKFWELGCDPGLETNGSGOYEISFVWMC 358  
Db 316 ERETAKRHNSQPRSPFTHAGGLPAINRMLPEKLGYYDEGIQIINGSGOYEISFVW 375  
Qy 359 RMEDIPCSRVGHIYRKVYVY---KVPAGVSLARULKRYAEVMDYAEYIYQRPERYHL 415  
Db 376 GIVVPPSHVGHVTRSHMPISFGKSGKPVISIMMRVYKTMDDYSKITIYTRBPQATNV 435  
Qy 416 SAGDAVQKRLSSLNCKSPKFMPTKIAMDLPKFYPVPEPAAWGEIRVNGTGLCADTK 475  
Db 436 NPGDISQLLRDLCKSPKFMWENVAVDLKSYPML-PENDVWGEARNPATGCLD-R 493  
Qy 476 HGAIGSPRLREGCYRGGEAANNMNVFTFWREDIRPQDQHTKKCFDAISHTSPVTL 535  
Db 494 MGGIPGEMGATGC-----HGYGQQLIRLVNQGQMGQGE-----MCLTA---NGIRI 537  
Qy 536 YDCHSMKG--NQLWKY-RKDKTLVHPVSGSCMDSQSDHRIEMNTCNSSLTQOMLPE 590  
Db 538 QANHCVTGVNGFWSYDRKTKQIHSQKQCIIVSSESEVTLQITCEDNEROKFVK 595

## RESULT 6

T42247  
polypeptide N-acetylgalactosaminyltransferase (EC 2.4.1.41) - Caenorhabditis elegans  
C/Species: Caenorhabditis elegans  
C/Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 09-Jul-2004  
C/Accession: T42247  
J. Hagen, F.K.; Nehrke, K.  
R. Biol. Chem. 273, 8268-8277, 1998  
A/Title: cDNA cloning and expression of a family of UDP-N-acetyl-Dgalactosamine:polypept  
A/Reference number: Z22126; MUID:98192620; PMID:9525933  
A/Accession: T42247  
A/Status: preliminary; translated from GB/EMBL/DBJ  
A/Molecule type: mRNA  
A/Residues: 1-624 <HAG>  
A/Cross-references: UNIPROT:O61393; EMBL:AF031837; NID:g3047192; PIDN:AAC13672.1; PID:g3  
C/Genetic8:  
A/Gene: gly-5  
C/Superfamily: polypeptide N-acetylgalactosaminyltransferase  
C/Keywords: glycosyltransferase; hexosyltransferase

Query Match 33.0%; Score 1083; DB 2; Length 624;  
Best Local Similarity 45.0%; Pred. No. 1.7e-79;  
Matches 226; Conservative 76; Mismatches 166; Indels 34; Gaps 13;

Qy 103 RVDAVYENGFNIVSDKISLNRSLPDIRHPCNSKRYLETPNTSIIIPHNMGWSSL 162  
Db 133 KYDGMALNNFNOYASDMISVHRTLPTNIDAECKTEKYNELPRTSVIICPHNEAMSVLL 192  
Qy 163 RTVSHVLRSPPELVAEIVLVDDPSDREHLKKPLEDYMLF-PSVRLRTKKEGLIRTR 221  
Db 193 RTVSHVLRTPDHLLEEVVLVDPSDMHTKRPLEEYVSQFGKVKILRMKREGLIRAR 252  
Qy 222 MLGASVATGVTITFLDSHCENNVMLPRLDIRIARNKTIYCPMIDVIDHDPRY-ETQA 280  
Db 253 LRGAVALVGEVLITLDSCHECMGMEPLDIRIKDPTTVVCPVIDVIDDTFETHSKA 312  
Qy 281 GDARGAFDWMYKRIPIPELOK--ADPSDPESPESPMAGGLFAVDRKFWELGSDYDPG 338  
Db 313 YFTSVGGFDKGLQNMHSIPERDRKNTRPIDPVRSPTMAGGLFSIDKEYEKGLTYDPG 372  
Qy 339 LEIWGSGOYEISFVWMCGRMEDIPCSRVGHIYRKVYVYKVPAGV--LAENLKRVAEW 397  
Db 509 EDIRPQHTKKECFDAISHTSPVTLYDCHSMKGNOLMKYRKDK-TLYHPVSGSCMDSD 567

Db 373 FDIWGENLELSEFKIMWCGGTLEIVPCSHVGHVFRKRSPYKMTGTVNLKRNIRLAEW 432  
Qy 398 MDEYAEYIYQRPERYHLASDAVAQKRLSSLNCKSPKFMPTKIAMDLPKFYPVPEPA 457  
Db 433 LDDYKTYTTERINNQLG-DIGDISSRKKLRBDLGCSFKYIYLDNI---YBELFVPGS--S 486  
Qy 458 AAMGEIRNVGTGLCADTKHAGLSPRLREG---CYRGGEAANNMNVFTFWREDIR 512  
Db 487 VAKGELNNAQTSQCLDS---AVGEVENKAITTPYCHEQSGNOYW-----MLSKDEIR 537  
Qy 513 PGDPQHTKKECFDAISHTSPVTLYDCHSMKGNOLMKYRKDK-TLYHPVSGSCMDSD 571  
Db 538 RDES-----CVDYAG--SDVMVFPCHGMKGNDEMRYNHDTGRLOHVAVSQKCLAMTXDGA 589  
Qy 572 RIFMNTGNPSSLTQOMLFETHN 593  
Db 590 KLENVACQYDDPYQHMFKKEYN 611

## RESULT 7

T42246  
polypeptide N-acetylgalactosaminyltransferase (EC 2.4.1.41) - Caenorhabditis elegans  
C/Species: Caenorhabditis elegans  
C/Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 09-Jul-2004  
C/Accession: T42246  
J. Hagen, F.K.; Nehrke, K.  
R. Biol. Chem. 273, 8268-8277, 1998  
A/Title: cDNA cloning and expression of a family of UDP-N-acetyl-Dgalactosamine:polypept  
A/Reference number: Z22126; MUID:98192620; PMID:9525933  
A/Accession: T42246  
A/Status: preliminary; translated from GB/EMBL/DBJ  
A/Molecule type: mRNA  
A/Residues: 1-626 <HAG>  
A/Cross-references: UNIPROT:O61392; EMBL:AF031836; NID:g3047192; PIDN:AAC13672.1; PID:g3  
C/Genetic8:  
A/Gene: gly-5  
C/Superfamily: polypeptide N-acetylgalactosaminyltransferase  
C/Keywords: glycosyltransferase; hexosyltransferase

Query Match 32.9%; Score 1079; DB 2; Length 626;  
Best Local Similarity 44.9%; Pred. No. 3.5e-78;  
Matches 227; Conservative 75; Mismatches 164; Indels 40; Gaps 14;

Qy 103 RVDAVYENGFNIVSDKISLNRSLPDIRHPCNSKRYLETPNTSIIIPHNMGWSSL 162  
Db 133 KYDGMALNNFNOYASDMISVHRTLPTNIDAECKTEKYNELPRTSVIICPHNEAMSVLL 192  
Qy 163 RTVSHVLRSPPELVAEIVLVDDPSDREHLKKPLEDYMLF-PSVRLRTKKEGLIRTR 221  
Db 193 RTVSHVLRTPDHLLEEVVLVDPSDMHTKRPLEEYVSQFGKVKILRMKREGLIRAR 252  
Qy 222 MLGASVATGVTITFLDSHCENNVMLPRLDIRIARNKTIYCPMIDVIDHDPRY-ETQA 280  
Db 253 LRGAVALVGEVLITLDSCHECMGMEPLDIRIKDPTTVVCPVIDVIDDTFETHSKA 312  
Qy 313 YFTSVGGFDKGLQNMHSIPERDRKNTRPIDPVRSPTMAGGLFSIDKEYEKGLTYDPG 372  
Qy 339 LEIWGSGOYEISFVWMCGRMEDIPCSRVGHIYRKVYVYKVPAGV--LAENLKRVAEW 397  
Db 373 FDIWGENLELSEFKIMWCGGTLEIVPCSHVGHVFRKRSPYKMTGTVNLKRNIRLAEW 432  
Qy 398 MDEYAEYIYQRPERYHLASDAVAQKRLSSLNCKSPKFMPTKIAMDLPKFYPVPEPA 457  
Db 433 LDDYKTYTTERINNQLG-DIGDISSRKKLRBDLGCSFKYIYLDNI---YBELFVPGS--S 486  
Qy 458 AAMGEIRNVGTGLCADTKHAGLSPRLREGCYRGGEAANNMNVFTFWREDIR 512  
Db 487 VAKGELNNAQTSQCLDS---AVGEVENKAITTPYCHEQSGNOYW-----MLSKDEIR 537  
Qy 509 EDIRPQHTKKECFDAISHTSPVTLYDCHSMKGNOLMKYRKDK-TLYHPVSGSCMDSD 567

Db 532 LS-KDGEIRDES-CVDYAG--SDVWVFPCHGKMGKNGEMRYNHDTRGLQHAVSQKLGMT 587  
QY 568 BSDHRIFMANTCNPSSLTQOMLFEBHTN 593  
Db 588 KDGAKLEMAVCQYDDPYQHWKFKKEYN 613

RESULT 8  
T42245  
probable polypeptide N-acetylglactosaminyltransferase (EC 2.4.1.41) - Caenorhabditis el  
C/Species: Caenorhabditis elegans  
C/Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 09-Jul-2004  
C/Accession: T42245  
R:Hagen, R.K.; Nehrke, K.  
J. Biol. Chem. 273, 8268-8277, 1998  
A>Title: cDNA cloning and expression of a family of UDP-N-acetyl-Dgalactosamine:Polyper  
A/Reference number: Z22126; MUID:98192620; PMID:9525933  
A/Accession: T42245  
A/Status: preliminary; translated from GB/EMBL/DBJ  
A/Molecule type: mRNA  
A/Residues: 1-623 <HAG>  
A/Cross-references: UNIPROT:O61391; EMBL:AF031835; NID:g3047190; PIDN:AAJ3671.1; PID:g3  
A/Genes: g3ly-5  
C/Superfamily: polypeptide N-acetylglactosaminyltransferase  
C/Keywords: glycosyltransferase; hexosyltransferase

Query Match 32.9%; Score 1078.5; DB 2; Length 623;  
Best Local Similarity 45.3%; Pred. No. 3.8e-79;  
Matches 226; Conservative 74; Mismatches 170; Indels 29; Gaps 13;

QY 103 RVQAIRENGFNIVYSKISLNSLPDIRHPNCKSKYLETLPNTSIIIPHNAGSSLL 162  
Db 133 KYDKGMLNNAFNOYASDMISVHRTLPTNIDAECKTEKYNELPTSTYIICFHNEAMSVLL 192  
QY 163 RTHSVNLRSPPELVAVIYVDDPSREHKKPLEDMALF-PSVRLTRTKKRGILRTR 221  
Db 193 RTHSVNLRSPPELVAVIYVDDPSREHKKPLEDMALF-PSVRLTRTKKRGILRTR 252  
QY 222 MCGASVATGVTITPLDSHCANVMNLPPLDRIARNKTIYCPMIDVIDHDDFRY-ETOA 280  
Db 253 LRGAATAVGTGLVTLDSHCMEGMEPLDRIARDPTTVCPVIDVIDNTPEYHNSKA 312  
QY 281 GDAMKAFDMWKYKRIPIPELOK--ADSPDPESPVMAAGLFAVDKRMFELGYDPG 338  
Db 313 YFTSVGDFDGLQFMNHSIPERDKNTRPIDPVRSPTMAGLFSIDKEFEKLGTYDPG 372  
QY 339 LETWGGQVYISPFVMMCGRMEDIPCSRYGHIYRKVPKVPAGVS-LARNLKRVAEWM 397  
Db 373 FDIWGGNLTLSFKIMWCGTLEIVPCSHVGHVRRKSPYKMTGAVNLKRNISIRLAEVW 432  
QY 398 MDEYAEIYORREYRHLASGVAVQKRLSSLNCKSPKWMFKIADMLPKFYPPVPPA 457  
Db 433 LDDYKTYVYRINNQLG-DFGDISSRKKREDLGCKSKFYKWLNI---YELFLPFGH--S 486  
QY 458 AAMGEIRNVG--TGLCADYTHGALGSLRLGECYRGGGAANNMQUFTTWRREDIRPD 515  
Db 487 VAKGEMNNAAGKQRQCIDYKPSG-GKTVGWYQCNQGNQYW-----MLSKGEIRDRS 539  
QY 516 PQHTKKKCFPAISHTSPVTLDYCHSMKGNOLMKYRKQ-TLYHPVSGSCMDCSBHRIF 574  
Db 540 S-----CVDYAG--SDVWVFPCHGKMGKNGEMRYNHDTRGLQHAVSQKLGMTKXAKLB 591  
QY 575 MNTCNPSLSTQOMLFEBHTN 593  
Db 592 MVACQYDDPYQHWKFKKEYN 610

RESULT 9  
A88515  
polypeptide N-acetylglactosaminyltransferase homolog - Caenorhabditis elegans  
C/Species: Caenorhabditis elegans  
C/Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 02-Nov-2001

C/Accession: A88515  
R:Anonymous, The C. elegans Sequencing Consortium.  
Science 282, 2012-2018, 1998  
A>Title: Genome sequence of the nematode C. elegans: a platform for investigating biolog  
A/Reference number: A75000; MUID:99069613; PMID:9851916  
A/Note: see webistes genome.wustl.edu/gsc/C. elegans/ and www.sanger.ac.uk/Projects/C. ele  
A/Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and  
A/Accession: A88515  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-563 <STO>  
A/Cross-references: GB:chr\_III; PIDN:AAA28224.1; PID:g289776; GSPDB:GN00021  
C/Superfamily: polypeptide N-acetylglactosaminyltransferase

Query Match 30.2%; Score 988.5; DB 2; Length 563;  
Best Local Similarity 40.5%; Pred. No. 6.5e-72;  
Matches 211; Conservative 87; Mismatches 182; Indels 41; Gaps 15;

QY 88 GNGEGRPYPMTDARVDQYRENGENIYVSDKISLNSLPDIRHPNC-NSKRYLET--L 144  
Db 61 GGGGTGTVPEDEKTKIEKRFLENQFNVASSEMISVNRTPDYRSDCRTSGNNLKTAGM 120  
QY 145 PNTSIIIPHNAGSSLLRTHSVNLRSPPELVAVIYVDDPSREHKKPLEDMALF 204  
Db 121 PNTSIIIPHNAGSSLLRTHSVNLRSPPELVAVIYVDDPSREHKKPLEDMALF 180  
QY 205 -SVRIITKRGEGILRTRMAGASVATGVTITPLDSHCANVMNLPPLDRIARNKTIYVC 263  
Db 181 IPIHVLNRSGLIRARLGSSEMAKGLIFLDANVAVYDMLRPLVSAVEDRKKVVA 240  
QY 264 PMIDVIDHDDFRYETQAGDAMKAFDMWKYKRIPIPE-ELQK--ADSPDPESPVMAAG 320  
Db 241 PIIIDVIDHDDFRYET--ASETTWGGFNMHLNFRWAVAKRELNRGSDRSMPIQTPIAGG 299  
QY 321 LFAVDKRMFELGYDPGLEINGEYQISFKYMMCGRMEDIPCSRYGHIYRKVPKVPAGVS 380  
Db 300 LFAIDKQPFYDIDSGYMGWGMGEMLEISFRWMMCGSLIHPCHSVGHVFRQTYTF 359  
QY 381 PAGVS--LAANLKRVAEWMDEYAEIYORREYRHLASGVAVQKRLSSLNCKSPKMF 438  
Db 360 PGSTAKYIHNAARTAEVWMDYKAPFYKVPAAKRVABQVSRKGLRETLOCKSKMY 419  
QY 439 MTKIAMDLPKRYPPVEPPA--AAMGEIRNVGTGLCADYTHGALGSLRLGECYRGGBA 496  
Db 420 LENI-----YEPALPLADFRSLGALVNRFTKCVDTNGKQDQAPGIDACHGAGNOA 472  
QY 497 WNNMQUVTFPTWRREDIRPDQHTKKKCFDAISHT---SEVTLDYCHSMKGNOLMKYRKD 552  
Db 473 W-----SLTGKGEIRSD-----LCLSS-GHYTOIGSELKERCSSKINVGHVFPD 519  
QY 553 K---TLYHPVSGSCMDCSBHRIFMNTCNPSLSTQOMLE 590  
Db 520 DQAGTILHKKTKGCV--TGADQVTVLDECGLGKQDQWQLE 558

RESULT 10  
T42243  
probable polypeptide N-acetylglactosaminyltransferase (EC 2.4.1.41) - Caenorhabditis el  
C/Species: Caenorhabditis elegans  
C/Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 09-Jul-2004  
C/Accession: T42243  
R:Hagen, R.K.; Nehrke, K.  
J. Biol. Chem. 273, 8268-8277, 1998  
A>Title: cDNA cloning and expression of a family of UDP-N-acetyl-Dgalactosamine:Polyper  
A/Reference number: Z22126; MUID:98192620; PMID:9525933  
A/Accession: T42243  
A/Status: preliminary; translated from GB/EMBL/DBJ  
A/Molecule type: mRNA  
A/Residues: 1-612 <HAG>  
A/Cross-references: UNIPROT:P34678; EMBL:AF031833; NID:g3047186; PIDN:AAJ3669.1; PID:g30  
R:Wilson, R.

submitted to the EMBL Data Library, May 1993  
A:Description: Sequence of the C. elegans coemid ZK688.

A:Reference number: S44913

A:Accession: S44913

A:Molecule type: DNA

A:Residues: 50-612 <MIL>

A:Cross-references: EMBL:L16621; NID:G289775; PIDN:AAA28224.1; PID:G289776

C:Genetics:

A:Gene: g1y-3

A:Insertions: 101/1; 229/3; 520/3; 570/3

C:Superfamily: polypeptide N-acetylglucosaminyltransferase

C:Keywords: glycosyltransferase; hexosyltransferase

Query Match 30.2%; Score 988.5; DB 2; Length 612;  
Best Local Similarity 40.5%; Pred. No. 7.4e-72;  
Matches 211; Conservative 87; Mismatches 182; Indels 41; Gaps 15;

Qy 88 GNGQGRPYPTMDERDQAVRENGFNIVYSDKISLRSLFDIHPNC-NSKRLER-L 144  
Db 110 GGGGTGVTPEDKKTIRKRELENOFNVASEMISVNTLPDYSDACRTSGNNLKTAGM 169

Qy 145 PNTSIIIPFNHSGSLRTVHSVLYNSPPELVAEIVLVDPFSREHLKKLEDMYALFP 204  
Db 170 PKSIIVFNHNAETTLRTIHSYINSFRLLEIILVDDKSDRDYLVKELDIYIKFP 229

Qy 205 -SVRLTKRREGLIRTRMLGASVATGDTVITFLDSHCANVNLPELDRJARRKTIYC 263  
Db 230 IPIHLVLENRSGILIRALITGSEMAKRIILFLDAHVEVTGWLPLEVSRVAEDRKRVVA 289

Qy 264 PMDIVIHDDPRYTQAGDMRGAPEMRYKRIPIRP-ELQK--ADSPDFSPVMAAG 320  
Db 290 PIIIDVIDDTPPEYLT-ASETTWGGFNMTLNFMYAVPGRGLNRRGSDRSMPIQPTIAGG 348

Qy 321 LEAVDRKMFELGSDYDGLIWMGEQYEIFSKVMCGRMEDIPCSRGIHYRYKVPYKV 380  
Db 349 LFIADKQFTYDYGDEMGWGENLEISFRVMCGSLIHCSCRGVHFRKQTPYTF 408

Qy 381 PAGVS--LARNLKEVAEYVMDYAEYIYORRPEYRHLISAGDAVAVQKLRSLNCKSPKMF 438  
Db 409 PGGAFAKVIHNAATAEVWDEYKAFYKQVPAARVAGVSRKRLRETLQCKSPKMY 468

Qy 439 MTKIAMDLPKRYPEVPEPA--AANGELRNVTGICADTGHGALSPRLBECRGREGAA 496  
Db 469 LENT-----YPRAPLPADFRSLGAIYNRTEKCVDTNGKDDQAPCIQACHGAGGQA 521

Qy 497 WNNQVFTFTWRBEDIRPDDPQHTKKFCDAISHT-----SPVLYDCHSMKGNQLMKYRKD 552  
Db 522 W-----SLTGKGEIRSD-----LCUSS-GHYVQIQLSELKRCRSVSKINXHVFPD 568

Qy 553 K---TLVHPVSGSCMDCSDHRIEMNTCNPSILTOQLFE 590  
Db 569 DQAGTLHKTKGKCV--TGADQRTVLTDCGGLGRDQWQMLE 607

#### RESULT 11

T42249 polypeptide N-acetylglucosaminyltransferase (EC 2.4.1.41) 6b - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 09-Jul-2004

C:Accession: T42249

R:Hagen, F.K.; Nehrkke, K.

J. Biol. Chem. 273, 8268-8277, 1998

A:Title: cDNA cloning and expression of a family of UDP-N-acetyl-D-glucosamine:Polypept

A:Reference number: Z22126; MUID:98192620; PMID:9525933

A:Accession: T42249

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-617 <HAG>

A:Cross-references: UNIPROT:O61395; EMBL:AF031839; NID:G3047198; PIDN:ANCI3675.1; PID:G3

C:Genetics:

A:Gene: g1y-6

C:Superfamily: polypeptide N-acetylglucosaminyltransferase

C:Keywords: glycosyltransferase; hexosyltransferase

Query Match 29.5%; Score 968; DB 2; Length 617;  
Best Local Similarity 39.3%; Pred. No. 3.4e-70;  
Matches 216; Conservative 96; Mismatches 162; Indels 76; Gaps 22;

Qy 73 DMHDKAIRDQAVGNGEGRPYPTMDARV--DQAVRENGFNIVYSDKISLRSLFDIR 131  
Db 97 DW-----GGGAGVSHLTPEQGLADSTFAVQFNLVSDGISVRSRLEIR 143

Qy 132 HPNCSKRYLETLPNTSIIIPFNHSGSLRTVHSVLYNSPPELVAEIVLVDPFSRREH 191  
Db 144 KPCSRNMTYDNLPTTSVILVHNEASTLIRTYWSYIDSPRELLEIILVDDFSDREF 203

Qy 192 LKRP-LEDYALPPS-VRIIRTKRREGLIRTRMLGASVATGDTVITFLDSHCANVNLFP 249  
Db 204 LRYETLTTLTKLPFTDIIKIRSKERVGLIRARMGAGDAOGDVLTLFDSHCCECTKWLPR 263

Qy 250 LDRIRANRRTIYOPMIDVIDHDDPRYTQAGDMRGAPEMRYKRIPIRP-ELQK--A 306  
Db 264 LTRIKLNRAVPCPVVIDIINDTFOYQ-KGIEFRGCFNMNLOFRMYGMPYAAKQHL 322

Qy 307 DPSPDFSPVMAAGLFAVDRKMFELGSDYDGLIWMGEQYEIFSKVMCGRMEDIPCS 366  
Db 323 DPTGPIESPMAAGLFNINNYFELGSDYDGMIDWGENLEMSFRIMCGGRVETLPCS 382

Qy 367 RVGHIYKRYVPYKVP--AGVSLARNLKEVAEYVMDYAEYIYORRPE-YRHLISAGDAV 422  
Db 383 HVGVFPRKSSPHDPFGKSGKVLNTNLRVAEYVMDWDMKYFYKIALQARHMSISIVSE 442

Qy 423 QKLRSLNCKSPFEMTKIAMDLPKRYPEVPEPAAMGHIRNVTGLC-----ADT--- 474  
Db 443 RVELRKLNCKSPFMYQANVPQD--HFLP--TPLDIFGKSN--SNYCTAFRPGDTPK 495

Qy 475 KHGALGSPRLLEGCVRGGAAMNMQVFTFTWRBEDIRPDDPQHTKKFCDAIS--HTS- 531  
Db 496 NHRLLGSP-----CTMG-----FDLWQMLYLTGRRIRRTDE-----HLCLSVQGLHTTS 540

Qy 532 --PVTLYDCHSMKGNQLMKYR-KDKTLYHPVSGSCMDCSDHRIF-----MNT 577  
Db 541 DWKIQLEKAGF--DTEYWDKPKIGRFQNRKTCGLASPD---IFDPTKDFENPPIYOK 595

Qy 578 CNPSILTOQM 587  
Db 596 CRSSNDROW 605

#### RESULT 12

I37405 polypeptide N-acetylglucosaminyltransferase (EC 2.4.1.41) - human

C:Species: Homo sapiens (man)

C:Date: 24-May-1996 #sequence\_revision 24-May-1996 #text\_change 09-Jul-2004

C:Accession: I37405

R:White, T.; Bennett, E.P.; Paul, E.; Takio, K.; Stensen, T.; Bonding, N.; Clausen, H.

J. Biol. Chem. 270, 24156-24165, 1995

A:Title: Purification and cDNA cloning of a human UDP-GalNAc:polypeptide N-Acetylglucto

A:Reference number: I37404; MUID:96025800; PMID:7552619

A:Accession: I37405

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-571 <RES>

A:Cross-references: UNIPROT:Q10471; EMBL:X85019; NID:G971460; PIDN:CAA59381.1; PID:G9714

C:Genetics:

A:Gene: GDB:GALNT2; GALNAC

A:Cross-references: GDB:696223

A:Map position: 16q24-16q24

C:Superfamily: polypeptide N-acetylglucosaminyltransferase

C:Keywords: glycosyltransferase; hexosyltransferase

Query Match 29.0%; Score 949; DB 2; Length 571;  
Best Local Similarity 33.5%; Pred. No. 1.1e-68;  
Matches 221; Conservative 103; Mismatches 170; Indels 166; Gaps 21;

Qy 1 MRRERKLLQAVLVALVALVLPVNGLMALYRERQDPTGSGAVALPAAGGSHSRQK 60

```

Db      1 MRRSRMLTLCFAFLWLVGIA-----YNY-----SGGSALAGAGGGA----- 39
Qy      61 KTFPLDGGQKLDKMDKXEARIRDAQAVNGEQ-----GR-PYPMTDARV----- 104
Db      40 -----GRK-EDNNEIDIPKIKDLDHSHNGEKAQSMETLPPGKVRMPDFNGQAVVGTM 91
Qy      105 -----DQAVRNGFNIVYSDKISLRSLPDI RHPNCSKRYLETLPNTSIIIPFNHGMSS 160
Db      92 VRSGQDPYANKFKFQVESDKLRMDRAIPDTHDCCQKQKQWVDLPATSVTVTFPNEARSA 151
Qy      161 LKRTVSHVLSRSPPELVAEIVLVDFSDREHLKKPLD--YMALEPSPRIATKKRGLI 218
Db      152 LKRTVSVLKKSPPHLLKEIILVDDYGNP-----EDGALLGKIKRYVRANRRRGLM 205
Qy      219 RTMRLGASVATGVITFLDHSCEANVNLPLDLRIANRRTIVCPMIDVIDHDFRYET 278
Db      206 RSRVRGADAAQAKVLTFLDSHCECHHMLERVAEDRTRVVSPIIDVINMDNFGYVG 265
Qy      279 QADAMKGAFTWEMYKRIPIPELQKA---DPSDPESPVMAGGLFVADRKMFWELGKY 335
Db      266 ASAD-LKGGFDWMLVFKMDYMTPEQRSRQGNPAPIKTPMIAGLFVMDKFYEELGKY 324
Qy      336 DPGLEWGEBOYEISFKVMCGRMEDIPCSRVGHIYRKVYVYKVP--AGVSLARNLKRV 393
Db      325 DMMMDVWGENLEISFVWQCGSLLEIIPCSRVGHVFRKQHPYTFPGSGTVFARNTTRA 384
Qy      394 AEVMDDEYAEIYQRRPEYRLASADVAVOKRLSSLNCKSPKFMPTKIAMDLKFPYVP 453
Db      385 AEVMDDEYKMFYAAVPSARVVPYGNIOSRLERLKKLSCKRPFKYLENV-----YPL 437
Qy      454 EPPAAMGEIRNVGTGLCATTKGALGSPRLBECVGRGEAAMNMQVFTFRREDIRP 513
Db      438 RVPDH-----QDIAGAL----- 450
Qy      514 GDPOHTKKCFEPAISHTSP--VTLYDCHSMKGNOLMKYRKDKLYH-----PVSGS 562
Db      451 --QGGTN--CLDTIGHFADGVGVYEGCHNAGNQMALTKESYKAMDLCITVVDRAFGS 506
Qy      563 CM---DCSESDHR-----TFNNTCPSSLTQOOWLP 589
Db      507 LTKLGGRENDSRQKWEIGNSKLRHVGSNLCIDSRATKSGLSVEVCGP-ALSQQMKF 565

RESULT 13
T42248
polyptide N-acetylglucosaminyltransferase (EC 2.4.1.41) - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 09-Jul-2004
C:Accession: T42248; T23138
R:Hagen, F.K.; Nehrke, K.
J. Biol. Chem. 273, 8268-8277, 1998
A:Title: cDNA cloning and expression of a family of UDP-N-acetyl-D-glucosamine:polypep
A:Reference number: Z22126; MUID:98192620; PMID:9525933
A:Accession: T42248
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-618 <HAG>
A:Cross-references: UNIPROT:O61394; EMBL:AF031838; NID:g3047296; PIDN:AA033674.1; PID:g3
R:Barlow, K.
submitted to the EMBL Data Library, June 1998
A:Reference number: Z19695
A:Accession: T23138
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-618 <WTL>
A:Cross-references: EMBL:AL024499; PIDN:CA019707.1; GSPDB:GN00021; CESP:H38K22.5
C:Genetics:
A:Gene: H38K22.5
A:Map position: 3
A:Insertions: 42/3; 64/3; 154/1; 202/1; 368/3; 430/1; 477/3; 513/3; 563/3; 604/3
A>Note: gly-6

```

```

C:Superfamily: polyptide N-acetylglucosaminyltransferase
C:Keywords: glycosyltransferase; hexosyltransferase
Query Match      28.9%; Score 948.5; DB 2; Length 618;
Best Local Similarity 38.3%; Pred. No. 1.3e-60;
Matches 211; Conservative 100; Mismatches 163; Indels 77; Gaps 21;

Qy      73 DMDHKAIRDAQVNGEGQRPYPMTDARV--DQAVRNGFNIVYSDKISLRSLPDIR 131
Db      97 DW-----GGGAGVSHLTPEQQLADSTAVVQFNLVSDGISTVRSLEIR 143
Qy      132 HPNCSKRYLETLPNTSIIIPFNHGMSSLLRTVSHVLSRSPPELVAEIVLVDFSDR 191
Db      144 KPSGRNMTYEDNLPFTSVIIIVHNKASTLIRTWASVINDSPKLLKEIILVDFSDREF 203
Qy      192 LKRP-LSDYALFPS-VRIITKXREGIITRMLGASVATGVITFLDHSCEANVNLPP 249
Db      204 LRYPTLDTLTKPLPTDIKIRSKERVGLIRAWMGQAQGDVLTFLDSHCECTKWLEP 263
Qy      250 LLDRIANRRTIVCPMIDVIDHDFRYETQAGDAMKGAFTWEMYKRIPIPELQK--A 306
Db      264 LTRIKLNRRAVPCPVIDIINDNTFOYQ-KGIEWFRGFMWNLQFRWYGMPTAKOHL 322
Qy      307 DPSDPESPVMAGGLFVADRKMFWELGKYDPGLEWGEBOYEISFKVMCGRMEDIPCS 366
Db      323 DTPPTISPTMAGGLFISNNRYFELDEIYDGMIDWGENLENSFRIMCGRVEILPCS 382
Qy      367 RVGHIYRKVYVYKVP--AGVSLARNLKRYAEVMDDEYAEIYQRRPE-YHLSAGDVA 422
Db      383 HGVHVFRRKSPHDPFGKSGVINTNLRLVAEVMDDMKHYFYXIAQAHMRSSIDVSE 442
Qy      423 QKRLSSLNCKSPKFMPTKIAMDLKFPYVPEPPAAMGEI-RNVGGLCATDKGALGS 481
Db      443 RVELRKKLNCKSPFKYQNVFOQ--HPLP--YPLDRFGMTSSNSVSC-----LAW 490
Qy      482 PLRLEG-----CVARGEAAMNMQVFTFRREDIRPQDPOHTKKCFEPAIS--HTS 531
Db      491 TLNSSGKITKTSYADCLK-----IFHKTQLWLYTGDRIRITDE-----HLCLSVQLLHTT 540
Qy      532 ---PVTLYDCHSMKGNOLMKYR-KDKTLYHPVSGSCMDCSESDHRIF-----MN 576
Db      541 SDWKIQLEKQAGF-DTEYWFKPKLGRFQNRKGLGLASPD---IFDPTDBFNPPIVQ 595
Qy      577 TCNPSLTQOW 587
Db      596 KCRSSNDRONW 606

RESULT 14
T42250
polyptide N-acetylglucosaminyltransferase (EC 2.4.1.41) 6c - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 09-Jul-2004
C:Accession: T42250
R:Hagen, F.K.; Nehrke, K.
J. Biol. Chem. 273, 8268-8277, 1998
A:Title: cDNA cloning and expression of a family of UDP-N-acetyl-D-glucosamine:Polytype
A:Reference number: Z22126; MUID:98192620; PMID:9525933
A:Accession: T42250
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-562 <HAG>
A:Cross-references: UNIPROT:O61396; EMBL:AF031840; NID:g3047200; PIDN:AA033676.1; PID:g3
C:Genetics:
A:Superfamily: polyptide N-acetylglucosaminyltransferase
C:Keywords: glycosyltransferase; hexosyltransferase
Query Match      28.6%; Score 938; DB 2; Length 562;
Best Local Similarity 47.5%; Pred. No. 8.1e-60;
Matches 182; Conservative 74; Mismatches 103; Indels 24; Gaps 8;

Qy      73 DMDHKAIRDAQVNGEGQRPYPMTDARV--DQAVRNGFNIVYSDKISLRSLPDIR 131

```

```

Db 97 -----GCGAGVSHLTPEQQLADSTFAVNOFNILVSDGISVRSLSPEIR 143
Qy 132 HPNNSKRYLETLENTSIIIPFNHEGWSLLRTVSHVLRSPPELVAVIYVNDPSREH 191
Db 144 KPSERNNTYDPNLTSTVSVIYVHNEAVSTLLRTVSVIDRSPKELKEIILVDDPSREF 203
Qy 192 LKRC-LBDYMA/LPFS-VYILRTKKREGILRTMLGASVATGVTFLDSCHEANVM/LP 249
Db 204 LRYTLDTLTKPLPTDIKIFSKERVGLIRARMGAGAGDVTFLDSCHECTKMLER 263
Qy 250 LLDRIANRKTIVCPMIDVIDHDFFRYETQAGDAMRGAFDWMYKRIPIPELOK--A 306
Db 264 LLTRIKLRKAVPCPVIDIINDNTFYQ-KGIEHMRGGFNNTLQFRWYGMPTAMAKQILL 322
Qy 307 DPSDPFSPVWAGGLFAVDRKMFELGSDYDGLIHWGEQVEISFKVMCCGRMEDIPCS 366
Db 323 DPTGPISPTWAGGLFISINRYFELGSDYDGMIDIMGEMLEMSFRIMQCGRVEILPCS 382
Qy 367 RVGIYRYKYVPYKVP---AGVSLARNLKRVAVNMDEYAYIYORRPE-YRHLASGVAV 422
Db 383 HGVGVFRKSSPHDPGKSSGKVLNTNLRVAVEMMDMKHYFKIAPQAHMRSIDYSE 442
Qy 423 QKKLRSSLCKSKFKFMTKIAMD 445
Db 443 RVELRKKLCKSKFKMYLQNVFOD 465

```

## RESULT 15

```

JC5247
polypeptide N-acetylglucosaminyltransferase (EC 2.4.1.41) T3 - mouse
N/Alternate names: protein-UDP acetylglucosaminyltransferase
C/Species: Mus musculus (house mouse)
C/Date: 13-Mar-1997 #sequence_revision 25-Apr-1997 #text_change 09-Jul-2004
C/Accession: JC5247
R/Zera, J.; Hagen, F. K.; Ten Hagen, K. G.; Van Whyckhuysse, B. C.; Tabak, L. A.
Biochem. Biophys. Res. Commun. 228, 38-44, 1996
A/Title: Cloning and expression of mouse UDP-GalNAc:polypeptide N-acetylglucosaminyl-
A/Reference number: JC5247; MUID:97069650; PMID:8912633
A/Accession: JC5247
A/Molecule type: mRNA
A/Residues: 1-633 <ZAR>
A/Cross-references: UNIPROT:P70419; GB:U70538; NID:91575722; PIDN:AAB09579.1; PID:915757
A/Experimental source: testis
C/Comment: This enzyme catalyzes the formation of N-acetyl-D-galactosamine alpha-O-Ser/T
ictating O-glycosylation of serine and threonine residues on an array of glycoproteins.
C/Superfamily: polypeptide N-acetylglucosaminyltransferase
C/Keywords: glycosyltransferase; hexosyltransferase

```

Query Match 28.5%; Score 934; DB 2; Length 633;  
 Best Local Similarity 39.2%; Pred. No. 2e-67;  
 Matches 206; Conservative 98; Mismatches 179; Indels 42; Gaps 18;

```

Qy 90 GEGRPYPMTDARVDAQYRENG-----FNIVSDKISLNSL-PDIRHPNCSKRY--L 141
Db 122 GASGKPFKITHLSPBEQKEKERTKCFNAPASDRISLHRLDGPDTAPPECIEQKFKRC 181
Qy 142 ETLNPTSGIIPFNHEGWSLLRTVSHVLRSPPELVAVIYVNDPSREH/LKPLLEDYMA 201
Db 182 PPLPTSVIYVHNEAVSTLLRTVSHVSVSSPALIKETIILVDDASVDYDLHEKLEIEIK 241
Qy 202 LFPSEVRIARTKREGILRTMLGASVATGVTFLDSCHEANVM/LP/LDRIANRKTII 261
Db 242 QFSIVKIVRQGERKGLIARLLGAAVATAEITLFLDACEGCEYGLLEPLARIANENTAV 301
Qy 262 VCPMIDVIDHDFFRYETQAGDAMRGAFDWMYKRIPIPELOKADPSDPFSPV 316
Db 302 VSPDIASIDLNTFERNKSPYGSNNRGNFDMWSLSFGWESLPDHEKQGRKDETYPIKPT 361
Qy 317 MAGGLFAVDRKMFELGSDYDGLIHWGEQVEISFKVMCCGRMEDIPCSRVGHYRKCV 376
Db 362 FAGGLFISIKYFEHIGSYDEEMELWGEMLEMSFRVWQCGGLIMPCSVVGHVFRKS 421

```

```

Qy 377 PYKVPAGVS-LARNLKRVAVNMDEYAYIYORRPE---YRHLASGVAVQKKLRSSLN 431
Db 422 PHTEPKGTQYIARNOVALAEVWMDYKEIFYRRTDAKTIYKQKSPDLSKRFPIKRLQ 481
Qy 432 CKSKFKMTKIAMDLPKFFYPVEPPAAKGEIRVGVGLCADT-KHGLASPLLECCVR 490
Db 482 CKNFTWYLTNTI---YPAAYVDP/LNPVIS-GYKSVGPEPLCLDVGENNQGGKPLLYTC-H 536
Qy 491 GRGEAANNMNOVFETFRREDIRPDPOHT--KKFCDPAISHTSPV--TLYDCHS--MKGN 544
Db 537 GLG-----GMQYFYSQAKRER-----HNTQKELCLHATGQVGVQKACVYKGHRTIAPGE 586
Qy 545 QLMRYRKDKTLVHPSGSCMDCSSESDHRIFMNTCPSSLTQOMLF 589
Db 587 QIVERRDQQLYNLPLFKWCLIS-SNGEHPNLY-PCDATD/LQKWIF 629

```

## RESULT 16

T31549

polypeptide N-acetylglucosaminyltransferase (EC 2.4.1.41) 9 - Caenorhabditis elegans

C/Species: Caenorhabditis elegans  
 C/Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 09-Jul-2004  
 C/Accession: T31549; T42253

R/Matthews, L.

submitted to the EMBL Data Library, October 1999

A/Reference number: 221043

A/Accession: T31549

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: DNA

A/Residues: 1-579 <WIL>

A/Cross-references: UNIPROT:Q9U2C4; EMBL:AL117202; PIDN:CAB57897.1; CESP:Y47D3A.23

A/Experimental source: clone Y47D3A

R/Hagen, F. K.; Nehrke, K.

J. Biol. Chem. 273, 8268-8277, 1998

A/Title: cDNA cloning and expression of a family of UDP-N-acetyl-D-galactosamine:polypep

A/Reference number: 222126; MUID:98192620; PMID:9525933

A/Accession: T42253

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: mRNA

A/Residues: 1-202, 'D', 204-579 <HAG>

A/Cross-references: EMBL:AF031643; NID:93047206; PIDN:AAC13679.1; PID:93047207

C/Genetics:

A/Gene: CESP:Y47D3A.23; gly-9

A/Introns: 45/1; 179/1; 343/3; 373/1; 453/3; 491/3; 533/3

C/Superfamily: polypeptide N-acetylglucosaminyltransferase

C/Keywords: glycosyltransferase; hexosyltransferase

Query Match 27.6%; Score 906; DB 2; Length 579;  
 Best Local Similarity 40.2%; Pred. No. 3.3e-65;  
 Matches 212; Conservative 88; Mismatches 182; Indels 46; Gaps 22;

```

Qy 84 AQRVNGEGGRPYMT--DARVDAQYRENG--NIYVSDKISLNSL-PDIRHPNCSKRY 140
Db 70 APRGPGKGPVVLVTGDAE-LQADMKKFMVNVHSDKLSLDVDFPRLQCKDKITX 128
Qy 141 -LETLNPTSGIIPFNHEGWSLLRTVSHVLRSPPELVAVIYVNDPSREH/LKPLLEDY 199
Db 129 DYALPKTSVILITDEAMTFLRTVSHVLRSPPELOEVIILDDNSKRGLEPDRDH 188
Qy 200 MALF-PSYRIARTKREGILRTMLGASVATGVTFLDSCHEANVM/LP/LDRIANR 258
Db 189 IKRGGKVRLLRKRVHRLIRAKLAGAEVAGDITVFLDSCHEANHGLLEPIVORISDER 248
Qy 259 KTIYCPMIDVIDHDFFRYETQAGD-AMRGAFDWMYKRIPIPELOK--ADSPDPES 314
Db 249 TAIYCPMIDISDNTLAH--GDWSLSTGFSVAHLHTWEGLSBEBQKRTKPTDIYRS 305
Qy 315 PVWAGGLFAVDRKMFELGSDYDGLIHWGEQVEISFKVMCCGRMEDIPCSRVGHYRK 374
Db 306 PTMAGGLAANREKFFEVGVGDSEMDIHWGSENLISFPAWMCSSIEFTPCSHGHIFRA 365
Qy 375 YVPYKVPAGVS---LARNLKRVAVNMDEYAYIYORRPEYRHLASGVAVQKKLRSSL 430

```

Db 366 GHPYNTGRNNKDVGTNSKRLAEVWMDYKRLYYVHREDLRTKVDGLTARHBLKRL 425  
 Qy 431 NCSSEFKMTKIMDLPEFPPVPPPAANGELRN--GTGLCADT--KRGALGSPRLR 466  
 Db 426 NCRPFKMTFLDNIA--KCKFI--MDQVAVAGALHTVSGTRMCTDTLQDRCKNSQLGVP 481  
 Qy 487 GCVRGGEAAMNQVFTFTRMBDIRPGDQHTKKFCFDAISHTSPVTLTDCSHMKG--N 544  
 Db 482 HC--QGGK---SSPOLMSLSEKGNLR-----RENTC--ASEENGNTRMCTC--SKAQFN 527  
 Qy 545 QLMKRYKDKTLVHPVSGSCMDCE---SDHRIFMNTCNPSSLTQOMLF 589  
 Db 528 ERWAY--ENKMIIRNLKSKCKMSTANLTKPGDVAIYVE--CDEDEHQKNMF 573

## RESULT 17

T42244

probable polypeptide N-acetylglactosaminyltransferase (EC 2.4.1.41) - Caenorhabditis elegans  
 C:Species: Caenorhabditis elegans  
 C>Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 09-Jul-2004  
 C/Accession: T42244  
 J:Hagen, F.K.; Nehrke, K.  
 J. Biol. Chem. 273, 8268-8277, 1998  
 A:Title: cDNA cloning and expression of a family of UDP-N-acetyl-Dgalactosamine:polypeptide  
 A:Reference number: Z22126; MUID:98192620; PMID:9525933  
 A/Accession: T42244  
 A/Status: preliminary; translated from GB/EMBL/DBJ  
 A/Molecule type: mRNA  
 A/Residues: 1-589 <HAG>  
 A/Cross-references: UNIPROT:O61390; EMBL:AF031894; NID:g3047188; PIDN:AAC13670.1; PID:g3  
 C:Genetics:

A:Gene: gly-4  
 C:Superfamily: polypeptide N-acetylglactosaminyltransferase  
 C:Keywords: glycosyltransferase; hexosyltransferase

Query Match 26.9%; Score 883; DB 2; Length 589;  
 Best Local Similarity 37.4%; Pred. No. 2.5e-63;  
 Matches 209; Conservative 80; Mismatches 188; Indels 82; Gaps 18;

Qy 56 HSRQK-----KTF---FLDGQGLKDMHDKAIRDQAVNGEGRPYMTDAERVDQ 106  
 Db 83 HERTKDVTKTDFVEKFLNKGK---WH-----QG-----ED 111  
 Qy 107 AYENGFNIVSDKISLNSLPDIRHPNCSKRYLET--LNTSTIIIPFNHSGSLIRTV 165  
 Db 112 KYRANFENQASALNPTFRKIPDSREPOCRDVYSKVGQPTTVIITYHNEASSLIRTV 171  
 Qy 166 HSYLNSPPELVAVIYLVDDPSDRHKKPLEDYMAFPSPVRLIRTKKRGILRTMILGA 225  
 Db 172 PSYFNOSPPELLEIVLVDDNSQDVEIGKEL---AQIQRIYLRNNQREGILRSRVKGA 227  
 Qy 226 SVATGVDITFLDSHCANVMPLPDLRIARNKTIYCPMIDVIDHDHDFEYEQADAMR 285  
 Db 228 QVAPAPVLTFLDSHIECNQKMLBELARIENPRAVAVPIIIDVINDFYVQASAD--LR 286  
 Qy 286 GADWEMTYRIRIPIPELQK---ADPSDPESPVMAGGLFAVDRKFWELGGIDPGLIEIV 342  
 Db 287 GGDWMLTVFWEPMNEQLRKERNAHPAPYRSPMAGGLFALSKEMFNEIGTYDLDMEVW 346  
 Qy 343 GGEQVEISFYVMWCGGEMEDIPCSRVGHIYRKYVPYVPAVGS--LARNIKRYVAEVMDE 400  
 Db 347 GGNLEISFVWQGGSLIETPCSRVGHVFRKHPYFPFGSGNVFOKNTRRRALEVWDE 406  
 Qy 401 YAEIYQRPPEYRHLASGDAVQKGLSSLNCKSPKFMPTKIAMDLPEFPPVPPAAAV 460  
 Db 407 YKAIYIKNVPSARFVNGDITDRAIRIDRLQCKSPKMYLENV-----YPLEIPKTP 459  
 Qy 461 GELRWGTC--LCADTKHAGLSPRLREGCYRGGEAAMNQVF---TFTWRREDIRPGDP 516  
 Db 460 GKSFQKIKGILCLDSVARKSEAPGLFGCHGTGNGQEM---VFDQTKTFKNAI----- 510  
 Qy 517 QHTKKECFDAISHT--SPVTLTDCSHMKGNQMLKRYRDKTLVHPVSGSCMDCE---SDH 571

Db 511 ---SGLDPSNTENKTVTMVCKENLRPDITM---VVEKNGMLTGQGCCLTVNQSGGDW 564  
 Qy 572 RIFMNTCNPSSLTQOMLF 590  
 Db 565 LIYGACHEINLNGAQRIWIF 583

## RESULT 18

T27397

hypothetical protein Y75B8A.9 - Caenorhabditis elegans  
 C:Species: Caenorhabditis elegans  
 C>Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004  
 C/Accession: T27397  
 J:Barlow, K.  
 Submitted to the EMBL Data Library, November 1998  
 A:Reference number: Z20361  
 A/Accession: T27397  
 A/Status: preliminary; translated from GB/EMBL/DBJ  
 A/Molecule type: DNA  
 A/Residues: 1-605 <WIL>  
 A/Cross-references: UNIPROT:Q9XW72; EMBL:AL033514; PIDN:CAA22098.1; CESP:Y75B8A.9  
 A/Experimental source: clone Y75B8A  
 C:Genetics:  
 A:Gene: CESP:Y75B8A.9  
 A:Interons: 78/3; 370/3; 391/2; 529/3; 565/3  
 C:Superfamily: polypeptide N-acetylglactosaminyltransferase

Query Match 24.4%; Score 801; DB 2; Length 605;  
 Best Local Similarity 31.9%; Pred. No. 1.1e-56;  
 Matches 187; Conservative 102; Mismatches 194; Indels 103; Gaps 18;

Qy 80 IRDDAQVNGEGGR-----PYMTDAERVD----- 105  
 Db 44 IQRHKEKLNNGQGRHDPDDDEGAKEQBEDAVEKONIAAPPLKSTFTPPRSKEIIDT 103  
 Qy 106 -----QAYENGFNIVSDKISLNSLPDIRHPNCSKRYLETLPNTSI 149  
 Db 104 DLGKINGRAEDLOVBYGKYKYPNGLSDRISRKIKIDSRARCSLSLTSLSLPAASI 163  
 Qy 150 IIFPNHSGSLIRTVHVSILNSPPELVAVIYLVDDPSD---REHLKPLEDYMALPFS 205  
 Db 164 VVCYFNBPSPVLIRMVNSIDFRTPKPHLHILLVDDSSSEWSNATDEAIKYREKHIIQWEK 223  
 Qy 206 VRLIRTKKRGILRTMILGASVATGDIYITFLDSHCANVMPLPDLRIARNKTIYCPM 265  
 Db 224 VKELKTDKNEGLIRAKIFGARBRANGELVFLDSHCENBEMLPPLDQIKQNRRTVCPI 283  
 Qy 266 IDVIDDDPFYEQADAMRGAFDWEYTYRIRIPIPELQKADS---DPESPVMAGGLF 322  
 Db 284 IDIIDLITMKY--VESPVCTGGVMMAMTFK--WDYFHRSTYFEDPMNTVNPILKSPMAGGLF 340  
 Qy 323 AVDRKFWELGGYDPSGLIEWGEOVEISFPKVMCGGEMEDIPCSRVGHIYRKYVPYVPA 382  
 Db 341 AIDKEYFPELIGSDDEGMDVGAENVEISVRLVCGGELLIMPCSRGHIFRRQRPYGIKT 400  
 Qy 383 GVSILANLRKVAEVMDEIAEYIYQRPPEYRHLIS--AGDAVQKGLSSLNCKSPKFMPTK 441  
 Db 401 D--SMGNKSVRLAVWMDVEIYENFEARPNYRPTFDGLIRSLISLRNLQCKPEKYLEN 459  
 Qy 442 IAMDLPKFPVPPPAANGELRN-----GTG--LCADTKHG--ALGSPRLR 484  
 Db 460 I---YBELDPDNTF-----NQLNNQILVAGKTLIMANGTCHLSLNSGGRLANGNRVE 511  
 Qy 485 LBCSVRGGEAAMNQVFTFTRREDIRPGDPQHTKKFCFDAISHTSPVTLTDCSHMKGN 544  
 Db 512 MRCKNH-----MERMQQWYKYSTNELR---PFGSSRMCLDSLRGSLVIL---CHNGAH 559  
 Qy 545 QLMKRYKDKTLVHPVSGSCMDCESDHRI--FMNTCNPSSLTQOMLF 589  
 Db 560 QMWQVSNAGRLY---SRSVNKCATGSDNDVSALSTLKFCSLANSFOF 602

## RESULT 19

T42252  
poly(amide N-acetylglucosaminyl)transferase (EC 2.4.1.41) 8 - *Caenorhabditis elegans*  
C/Species: *Caenorhabditis elegans*  
C/Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 09-Jul-2004  
C/Accession: T42252; T27290  
R/Hagen, F.K.; Nehrke, K.  
J. Biol. Chem. 273, 8268-8277, 1998  
A/Title: cDNA cloning and expression of a family of UDP-N-acetyl-D-glucosamine:poly(amide N-acetylglucosaminyl)transferase  
A/Reference number: 222126; MUID:98192620; PMID:9525933  
A/Accession: T42252  
A/Status: preliminary; translated from GB/EMBL/DBJ  
A/Molecule type: mRNA  
A/Residues: 1-421 <HAG>  
A/Cross-references: UNIPROT:Q45293; EMBL:AF031842; NID:G3047204; PIDN:AA13678.1; PID:G3  
R/Matthews, L.  
submitted to the EMBL Data Library, October 1998  
A/Reference number: Z20338  
A/Accession: T27250  
A/Status: preliminary; translated from GB/EMBL/DBJ  
A/Molecule type: DNA  
A/Residues: 1-421 <MTL>  
A/Cross-references: EMBL:AL032622; PIDN:CAA21500.1; GSPDB:GN00021; CBSP:Y66A7A.6  
A/Experimental source: clone Y66A7A  
C/Genetics:  
A/Gene: gly-8; Y66A7A.6  
A/Map position: 3  
A/Intons: 36/2; 62/3; 97/2; 181/3; 226/2; 279/2; 318/3; 369/3  
C/Keywords: glycosyltransferase; hexosyltransferase

Query Match 19.3%; Score 632; DB 2; Length 421;  
Best Local Similarity 34.0%; Pred. No. 3.4e-43;  
Matches 138; Conservative 78; Mismatches 146; Indels 44; Gaps 10;

QY 83 DAGVNGGEQGRPY-----PMDAERVDQAYRN-----GPRIT 116  
DB 21 EAELTPKCEHVDPEYENLEGMLDKPLTE-RKCNHTLKENLTBASKESEMGKSPAPDAL 79  
QY 117 VSDKISLNRSLPDRHPNCSKRYLETLPNTSIIIPFNEGMSLLRTVSVLNRSPPEL 176  
DB 80 SESELTGPRNVRKQAHKCESEK-DAVSSTSVVHHNMLSTLIMINIIETPSL 138  
QY 177 VAEIVLVDDPSDRH-LKKPLEDY---MALPPSVRIITKREGLIRTRMGASVATGDV 232  
DB 139 LKEIVLVDDSEEDHVLTKHLEKFAKIKGLEDKLIKSEYRQGLIRAKVHASRLATGEV 198  
QY 233 ITFLDSHCEANVMPLPDLIRANRKTIVCPMTDIVDHDPRYETQAGDAMRGAFDDEM 292  
DB 199 IVEFDSHCEVAERWLEPLQPIKEDPKSIVLVVDLNPVSFDSPSM--VAKSGFDMGF 256  
QY 293 YKKRIPIPELQKADPSD--PFESFVMAAGLFAVDKMFELGVDPLGELIIGSGOYEIS 350  
DB 257 TFKKIYLPWEYETFPENNVRKFPNSPAMRGGLAMRKEVFLGVEDMDMEIIGSENIELS 316  
QY 351 FKVMCGRMEDIPCSRVRGHIYRKVYRYPKAVGSLA-RNLKRAVEMDEYAEIYQRR 409  
DB 317 LKAMLCGRVVAVACSRGVHFRMRPYTSKRGMDTALYNNAVAKTWLGEYSKFAVK 376  
QY 410 PEYHLSAGDVAVOKKRLSSLNCSFKFMKTIAMDLKPYRPPVER 455  
DB 377 PRGAKMVFGLDTEPMQVDRLLCKDMKFIENV-----YPELEP 415

RESULT 20  
H97186  
glycosyltransferase domain containing protein [imported] - *Clostridium acetobutylicum*  
C/Species: *Clostridium acetobutylicum*  
C/Date: 14-Sep-2001 #sequence\_revision 14-Sep-2001 #text\_change 09-Jul-2004  
C/Accession: H97186  
R/Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee, J.; Bacteriol. 183, 4823-4838, 2001  
A/Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium *Clostridium acetobutylicum* ATCC 25690  
A/Reference number: A96900; MUID:21359325; PMID:21359325

A/Accession: H97186  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-1044 <RUR>  
A/Cross-references: UNIPROT:Q97GN9; GB:AE001437; PIDN:AAK80283.1; PID:G15025335; GSPDB:GB  
A/Experimental source: *Clostridium acetobutylicum* ATCC824  
C/Genetics:  
A/Gene: CAC2327

Query Match 4.5%; Score 147; DB 2; Length 1044;  
Best Local Similarity 20.6%; Pred. No. 0.0025;  
Matches 48; Conservative 55; Mismatches 108; Indels 22; Gaps 8;

QY 141 LETLPTSIIPFNFNEGMSLLRTVSHVSNRSPPELVAVIVLVDPS---DREHLKKPLR 197  
DB 1 MNSVPRKSFIIIVNVR-GLQHLKNCFSLSKULSYSDKIEIVIVDNGSKDSVEFLKKN-- 57  
QY 198 DYMLPPSVRIITKREGLIRTRMGASVATGVDITFLDSHCEANVMPLPDLIRAR- 256  
DB 58 ----YPAVVIIRKDSNEGFAPKPDNDAAKTAEGEYLA LINDMKLDKQWINDMEETLENC 112  
QY 257 NRKTIVCPMDIVDHDPRYETQAGDAMRGAFDDEMYYKRIPIPELQKADPSDPFESPV 316  
DB 113 NDNSIVCAGSKIVWDDSKLDFAGSVSPAGYGYQYDGM----DIKNAKKTNEDRDI 167  
QY 317 M--AGLFAVDKRMFELGVDPLGELIIGSGOYEISFKVMCGRMEDIPCSR 367  
DB 168 LFAAGSGMLRKQVFIIGGFDKDYFAV-YEDVDLGRMLVGLYKFAF--CSK 217

RESULT 21  
D87531  
glycosyl transferase family protein CC2277 [imported] - *Caulobacter crescentus*  
C/Species: *Caulobacter crescentus*  
C/Date: 20-Apr-2001 #sequence\_revision 20-Apr-2001 #text\_change 09-Jul-2004  
C/Accession: D87531  
R/Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.I.  
B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolton  
n, J.; Brmlaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.  
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001  
A/Title: Complete Genome Sequence of *Caulobacter crescentus*.  
A/Reference number: A87249; MUID:21173698; PMID:11259647  
A/Accession: D87531  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-306 <STO>  
A/Cross-references: UNIPROT:Q9A618; GB:AE005673; NID:G13423792; PIDN:AAK24248.1; GSPDB:GB  
C/Genetics:  
A/Gene: CC2277  
C/Superfamily: Rhizobium succinoglycan biosynthesis glycosyltransferase

Query Match 4.2%; Score 139; DB 2; Length 306;  
Best Local Similarity 25.4%; Pred. No. 0.0019;  
Matches 57; Conservative 36; Mismatches 91; Indels 40; Gaps 10;

QY 144 LPTSIIPFNFNFNEGMSLLRTVSHVSNRSPPELVAVIVLVDPSDREHLKKPLEDYMA 201  
DB 1 MARTVMIIPQRRRDGLAAVARSVFGQVDFPAL--ELVIYD--NQVSAKAVADALR 56  
QY 202 L-PPSVRIITKREGLIRTRMGASVATGVDITFLDSHCEANVMPLPDLIRANRKT 260  
DB 57 KGAPCPVIYVEKRPVAFANAGMARASGFIIFLDDEBAPSGWLAALAAOERYDAD 116  
QY 261 IYCPMI-----DIVDDDF--RYETQAGDAMRGAFD-----EMYYKRIPIPELQKAD 307  
DB 117 VVEPVARAPAHIDQRDDIERFSRIGPQACVIVDHYGCGSLIRSLP-----D 170  
QY 308 PSDPF-----ESPVMAAGLFAVDKRMFELGVDPLGELI 342  
DB 171 PVAFAVARNFIDGEDDLRLHMGAAKRRAME-----PAAWV 209

RESULT 22

H87306  
glycosyl transferase family protein CC0465 [imported] - *Caulobacter crescentus*  
C:Species: *Caulobacter crescentus*  
C:Date: 20-Apr-2001 #sequence\_revision 20-Apr-2001 #text\_change 09-Jul-2004  
C:Accession: H87306  
R:Nalmeru, M.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.  
B.; Lab, M.T.; Deboy, R.T.; Dodson, R.U.; Durkin, A.S.; Gwinn, M.L.; Hatt, D.H.; Kolonin,  
n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.,  
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001  
A:Title: Complete Genome Sequence of *Caulobacter crescentus*.  
A:Reference number: A87249; MUID:21173698; PMID:11595647  
A:Accession: H87306  
A:Status: Preliminary  
A:Molecule type: DNA  
A:Residues: 1-308 <STO>  
C:Cross-references: UNIPROT:Q9AAK7; GB:AE005673; NID:gl3421638; PIDN:AAK22452.1; GSPDB:G  
C:Genetic:8;  
A:Gene: CC0465

**RESULT 25**

**D87506**

glycosyl transferase family protein CC2077 [imported] - *Caulobacter crescentus*

C:\Date: 20-Apr-2001 #bsequence\_20-Apr-2001 #text\_change 09-Jul-2004

C;Accession: D87506

C;Accession: D87506

R; Niernman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.

B.; Laub, M.T.; Deboy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Hatt, D.H.; Kolon  
T.; Ermojeva M.; White O.; Salzberg S.L.; Shavrylo I.; Venter J.C.; Fraser C M.

Proc. Natl. Acad. Sci. U.S.A. 98: 4136-4141, 2001

A:Title: Complete Genome Sequence of *Caulobacter crescentus*.  
FIOC: NAL1.ACau: SCI: 0:3.A: 30, 4138-4141, 2001

A:Reference number: A87249; MUID:21173698; PMID:11259647

A;Accession: D87506

A;Status: prelimina

A;Molecule type: DNA

A;Residues: 1-318 <ST

A;C:Cross-references: UNL

C/Gene: CC2077

11/07/2013 12:00:00 PM

Query Match

Best Local Similarity 24.7%; Pred. No. 0.0039;

Matches 63; Conservative 39; Mismatches 102; Indels 51; Gaps 10;

143 ТИ ДНТЕСТТТДЕНУЕОИОСИТДТНУЕУТ НРСБДБЕИ УАЕТИТ УНДЕСДБФУТ ККДТ ЕДУМАТ. 202

143 TLENTSII PFHNEGWSLLRIVHSV LNRSPP ELVAEIVLVDDFS DREHLKKPLEJEDYMAL 202

[illegible][illegible]

203 FPSVRLRTKKREGILRTMLGASVATGDVITFLDSHC EANNVWLPP---LDRIARNR 258

Db 72 EPRVRLQGLNIGFARAANLGATAKGEDEVFLNPDA-----FLTPGAIAALREARDR 126

[illegible]

259 KTI VCPMIDVIDHDDFRYEITQAGDAMRGAFD-----WEMYYKRILP 298

Db 127 P S P C V G A R V E N T D G --- T E O R G G R G E I T P V T T L I S I S K I S A T L P P L R R F E I H R E G E P 1822

DD 127 F0FCVGDHVENIDG-- --LEGNOKNGBE1FVI1LHSDOLWDBAI DFFLNKFBZHNDEBF 102

Search completed: November 22, 2004, 13:44:12  
Job time : 25 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 22, 2004, 13:38:22 / Search time 67 Seconds

(without alignments)  
5178.372 Million cell updates/sec

Title: US-10-001-851-2

Perfect score: 3278  
Sequence: 1 MRKKEKRLQALVALVALVLAALV.....TQOMLPENHTSYLAEKRN 603

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1825181 seqs, 575374646 residues

Total number of hits satisfying chosen parameters: 1825181

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

1: uniprot\_sprot.\*  
2: uniprot\_trembl.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3278	100.0	603	1	GLI0_HUMAN
2	3157	96.3	603	1	GLI0_RAT
3	3138	95.7	634	2	BAD21A05
4	3137	95.7	603	1	GLI0_MOUSE
5	3137	95.7	603	2	AAH60617
6	1404	42.8	644	2	Q700B9
7	1392.5	42.5	622	1	GLI0_CAEEL
8	1360	41.5	666	1	GLT6_DROME
9	1360	41.5	666	2	AAQ56703
10	1341.5	40.9	599	2	Q7PZM5
11	1247	38.0	1003	2	Q70DR0
12	1186	36.2	659	1	GLT4_DROME
13	1186	36.2	659	2	AAQ56701
14	1172.5	35.8	650	1	GLT9_DROME
15	1159	35.4	645	2	Q7Q815
16	1144	34.9	518	2	Q7Q048
17	1125	34.3	559	1	GLT1_MOUSE
18	1117	34.1	559	1	GLT1_PIG
19	1116	34.0	559	1	GLT1_RAT
20	1115	34.0	559	1	GLT1_BOVIN
21	1115	34.0	559	1	GLT1_MOUSE
22	1113	34.0	601	1	GLT7_CAEEL
23	1112.5	33.9	556	2	O6GMS1
24	1111.5	33.9	617	2	AAQ56702
25	1109	33.8	556	1	GLI3_MOUSE
26	1109	33.8	556	1	GLI3_RAT
27	1109	33.8	556	1	AAQ75749
28	1106.5	33.8	630	1	GLT5_DROME
29	1106	33.7	556	1	GLI3_HUMAN
30	1087	33.2	559	2	O6PAB1
31	1087	33.2	559	2	AAH60419

32	1083	33.0	626	1	GLT5_CAEEL	Q95211 caenorhabd
33	1083	33.0	653	2	O6NS24	O6NS24 xenopus lae
34	1083	33.0	653	2	AAH70527	AAH70527 xenopus lae
35	1053	32.1	653	2	Q6N821	Q6N821 xenopus tro
36	1053	32.1	653	2	AAH67317	AAH67317 xenopus t
37	1050	32.0	657	1	GLT7_HUMAN	O6S682 homo sapien
38	1049	32.0	930	1	GLT5_MOUSE	O8C102 m polypepti
39	1047	31.9	657	1	GLT7_MOUSE	O80V60 mus musculi
40	1040	31.7	657	1	GLT7_RAT	Q97065 rattus norv
41	1040	31.7	930	1	GLT5_RAT	O88422 r polypepti
42	1037	31.6	606	2	Q7QEH0	Q7QEH0 anopheles g
43	1036	31.6	578	1	GLT4_HUMAN	O8A440 h polypepti
44	1035.5	31.6	591	2	GLT7_DROME	O8M448 drosophila
45	1035.5	31.6	591	2	AAQ56704	AAQ56704 drosophil
46	1035	31.6	582	2	O6IR91	O6IR91 xenopus lae
47	1035	31.6	582	2	AAH71009	AAH71009 xenopus l
48	1033	31.5	578	1	GLT4_MOUSE	O08832 m polypepti
49	1033	31.5	578	2	AAH57882	AAH57882 mus muscu
50	1013	30.9	940	1	GLT5_HUMAN	Q72769 h polypepti
51	999.5	30.5	558	1	GLT1_HUMAN	O8A428 h putative
52	990.5	30.2	576	1	GLI2_MOUSE	Q8BGE3 m polypepti
53	990.5	30.2	576	2	AAH56425	AAH56425 mus muscu
54	989	30.2	563	2	Q6DJR8	O6DJR8 xenopus tro
55	988.5	30.2	581	1	GLT3_CAEEL	P34678 caenorhabd
56	986.5	30.1	581	1	GLI2_HUMAN	O81XK2 h polypepti
57	977.5	29.8	558	1	GLT1_MOUSE	Q9J161 m putative
58	971	29.6	552	1	GLI4_HUMAN	Q96E11 h polypepti
59	971	29.6	552	2	AAO89118	AAQ89118 homo sapi
60	965	29.4	550	1	GLI4_MOUSE	Q8BVG5 m polypepti
61	964	29.4	620	2	Q7Q456	Q7Q456 anopheles g
62	957.5	29.2	608	2	GLI1_MOUSE	Q92118 m polypepti
63	957.5	29.2	608	2	AAQ06668	AAQ06668 mus muscu
64	954.5	29.1	608	1	GLI1_HUMAN	Q8NCW6 h polypepti
65	954.5	29.1	608	1	GLI1_RAT	Q8P6V1 r polypepti
66	954.5	29.1	608	2	AAH62004	AAH62004 rattus no
67	953.5	29.1	590	2	AAH64620	AAH64620 drosophil
68	953.5	29.1	633	1	GLT2_DROME	O6WV19 drosophila
69	949	29.0	571	1	GLT2_HUMAN	Q10471 h polypepti
70	948.5	28.9	618	1	GLT6_CAEEL	O61394 caenorhabd
71	947.5	28.9	615	2	AAQ56700	AAQ56700 drosophil
72	942.5	28.8	570	1	GLT2_MOUSE	Q8PB93 m polypepti
73	942.5	28.8	570	2	AAH59818	AAH59818 mus muscu
74	935	28.5	633	1	GLT3_MOUSE	P70419 m polypepti
75	926	28.2	639	1	GLT2_HUMAN	Q8N311 h polypepti
76	926	28.2	639	2	AAQ88808	AAQ88808 homo sapi
77	925.5	28.2	622	1	GLT6_MOUSE	Q8C747 m polypepti
78	925	28.2	622	1	GLT6_MOUSE	Q8C14 h polypepti
79	923.5	28.2	667	1	GLT8_DROME	Q9VU16 drosophila
80	920	28.1	667	1	GLT3_DROME	Q9Y117 drosophila
81	916.5	28.0	601	2	AAQ56699	AAQ56699 drosophil
82	916	27.9	633	1	GLT3_HUMAN	Q14435 h polypepti
83	913.5	27.9	601	1	GLT1_DROME	O6WV20 drosophila
84	906	27.6	579	1	GLT9_CAEEL	Q9U2C4 caenorhabd
85	905	27.6	648	1	Q7Q8A5	Q7Q8A5 anopheles g
86	898	27.4	630	1	GLI0_DROME	Q81A43 drosophila
87	896	27.3	619	2	Q6N801	Q6N801 brachydanio
88	896	27.3	619	2	AAH67340	AAH67340 brachydan
89	896	27.3	627	2	Q7QIK5	Q7QIK5 anopheles g
90	883	26.9	589	1	GLT4_CAEEL	O81136 caenorhabd
91	873.5	26.6	638	1	GLI2_MOUSE	Q9D288 m polypepti
92	862	26.3	606	2	Q7Q0T6	Q7Q0T6 anopheles g
93	818.5	25.0	598	1	GLT3_HUMAN	O61824 h putative
94	818.5	25.0	598	2	AAH67524	AAH67524 homo sapi
95	818.5	25.0	598	2	AAH67525	AAH67525 homo sapi
96	818.5	25.0	598	2	AAH69624	AAH69624 homo sapi
97	818.5	25.0	598	2	AAH69628	AAH69628 homo sapi
98	818.5	25.0	598	2	AAH69636	AAH69636 homo sapi
99	818.5	25.0	598	2	AAH69645	AAH69645 homo sapi
100	818.5	25.0	598	2	AAH69997	AAH69997 homo sapi

## ALIGNMENTS

RESULT 1  
GL10\_HUMAN STANDARD, PRT, 603 AA.  
ID Q86SR1; Q86VP8; Q81XJ2; Q8TE32; Q961V2; Q9H8E1; Q9Y4M4;  
AC Q86SR1; Q86VP8; Q81XJ2; Q8TE32; Q961V2; Q9H8E1; Q9Y4M4;  
DT 01-OCT-2004 (Rel. 45, Created)  
DT 01-OCT-2004 (Rel. 45, Last sequence update)  
DT 01-OCT-2004 (Rel. 45, Last annotation update)  
DE Polypeptide N-acetylglucosaminyltransferase 10 (EC 2.4.1.41)  
DE (Protein-UDP acetylglucosaminyltransferase 10) (UDP-  
GlcNAc:polypeptide N-acetylglucosaminyltransferase 10) (Polypeptide  
GlcNAc transferase 10) (GlcNAc-T10) (pp-GalNAcTase 10).  
GN Name=GLNNT10;  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A. (ISOFORM 1), ENZYME ACTIVITY, AND TISSUE  
SPECIFICITY.  
RC TISSUE=Colon cancer;  
RX MEDLINE=22304871; PubMed=12417297;  
RA Cheng L., Tachibana K., Zhang Y., Guo J.-M., Tachibana K.K.,  
RA Kameyama A., Wang H., Hiruma T., Iwasaki H., Togayachi A., Kudo T.,  
RA Narimatsu H.;  
RT "Characterization of a novel human UDP-GlcNAc transferase, pp-GlcNAc-  
T10.";  
RL FEMS Lett. 531:115-121(2002).  
RN [2]  
RP SEQUENCE FROM N.A. (ISOFORM 1).  
RA Bennett E.P.;  
RL Submitted (AUG-2002) to the EMBL/Genbank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A. (ISOFORM 4).  
RC TISSUE=Uterus;  
RA Ansoorge W., Winkler U., Mewes H.-W., Gaessenhuber J., Wiemann S.;  
RL Submitted (JUN-1999) to the EMBL/Genbank/DBJ databases.  
RN [4]  
RP SEQUENCE FROM N.A. (ISOFORM 5), AND SEQUENCE OF 191-603 FROM N.A.  
(ISOFORM 1).  
RC TISSUE=Kidney, Skin, and Testis;  
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;  
RA Straubeberg R.L., Feinold E.A., Grouse L., Sherman C.M., Schuler G.D.,  
RA Klausner R.D., Collins F.S., Wagner L., Sherman C.M., Schuler G.D.,  
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heideh F.,  
RA Dichtenko L., Maruina K., Farmer A.A., Rubin G.W., Hong L.,  
RA Stepien M., Soares M.B., Bonaldi M.F., Casavant T.L., Scheetz T.E.,  
RA Brownstein M.J., Ueda T.B., Toshiyuki S., Carninci P., Prange C.,  
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,  
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Halyk S.W.,  
RA Villalón D.K., Wozny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,  
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
RA Butterfield A.S.N., Krzywinski M.T., Skalek U., Smalhus D.E.,  
RA Schermer A., Schein J.E., Jones S.J.M., Marra M.A.;  
RT "Generation and initial analysis of more than 15,000 full-length human  
RT and mouse cDNA sequences.";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
RN [5]  
RP SEQUENCE OF 6-366 FROM N.A. (ISOFORM 3), AND SEQUENCE OF 117-603 FROM  
RP N.A. (ISOFORM 2).  
RC TISSUE=Placenta, and Spleen;  
RX PubMed=14702039; DOI=10.1038/ng1285;  
RA Ota T., Suzuki Y., Nishikawa T., Otsuki T., Sugiyama T., Irie R.,  
RA Makamatsu A., Hayashi K., Sato H., Nagai K., Kimura K., Makita H.,  
RA Sekine M., Ohayashi M., Nishi T., Shibahara T., Tanaka T., Ishii S.,  
RA Yamamoto J.-I., Saito K., Kawai Y., Isono Y., Nakamura Y.,  
RA Nagahara K., Murakami K., Yasuda T., Iwayanagi T., Wagatsuma M.,  
RA Shiratori A., Sudo H., Hosoiri T., Kaku Y., Kodaira H., Kondo H.,

RA Sugawara M., Takahashi M., Kanda K., Yokoi T., Furuya T., Kikkawa E.,  
RA Omura Y., Abe K., Kamihara K., Katsuta N., Sato K., Tanikawa M.,  
RA Yamazaki M., Nimomiya K., Ishibashi T., Yamashita H., Murakawa K.,  
RA Fujimori K., Tanai H., Kimata M., Matanabe M., Hirao S., Chiba Y.,  
RA Ishida S., Ono Y., Takiguchi S., Matanabe S., Yoshida M., Horiuchi T.,  
RA Kusano Y., Kanehori K., Takahashi-Fujii A., Hara H., Tanase T.-O.,  
RA Imase N., Mueasahino K., Yuuki H., Oshima A., Sasaki K., Arita M.,  
RA Imose N., Mueasahino K., Yuuki H., Oshima A., Sasaki K., Arita M.,  
RA Yoshikawa Y., Matsunawa H., Ichihara T., Shiohara N., Sano S.,  
RA Moriya S., Momiyama H., Satoh N., Takami S., Tereshima Y., Suzuki O.,  
RA Nakagawa S., Senoh A., Mizoguchi H., Goto Y., Shimizu F., Wakebe H.,  
RA Hishigaki H., Matanabe T., Sugiyama A., Takemoto M., Kawakami B.,  
RA Yamazaki M., Matanabe K., Kumagai A., Itakura S., Fukuzumi Y.,  
RA Fujimori Y., Komiyama M., Tashiro H., Tanigami A., Fujiwara T.,  
RA Ono T., Yamada K., Fujii Y., Ozaki K., Hiro M., Ohnori Y.,  
RA Kawabata A., Hikiya T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,  
RA Okitani R., Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T.,  
RA Matsunura K., Nakajima Y., Mizuno T., Morinaga M., Sasaki M.,  
RA Togaishi T., Oyama M., Hata H., Matanabe M., Komatsu T.,  
RA Mizushima-Sugano J., Satoh T., Shirai Y., Takahashi Y., Nakagawa K.,  
RA Okumura K., Nagase T., Nomura N., Kikuchi H., Maeno Y., Yamashita R.,  
RA Nakai K., Yada T., Nakamura Y., Ohara O., Isogai T., Sugano S.;  
RT "Complete sequencing and characterization of 21,243 full-length human  
RT cDNAs.";  
RL Nat. Genet. 36:40-45(2004).  
CC -1- FUNCTION: Catalyzes the initial reaction in O-linked  
CC oligosaccharide biosynthesis, the transfer of an N-acetyl-D-  
CC galactosamine residue to a serine or threonine residue on the  
CC protein receptor. Has activity toward Muc5Ac and EA2 peptide  
CC substrates.  
CC -1- CATALYTIC ACTIVITY: UDP-N-acetyl-D-galactosamine + polypeptide =  
CC UDP + N-acetyl-D-galactosaminyl-polypeptide.  
CC -1- COFACTOR: Manganese and calcium (By similarity).  
CC -1- PATHWAY: Glycosylation.  
CC -1- SUBCELLULAR LOCATION: Type II membrane protein. Golgi (By  
CC similarity).  
CC -1- ALTERNATIVE PRODUCTS:  
CC Event=Alternative splicing; Named Isoforms=5;  
CC Name=1;  
CC IsoId=Q86SR1-1; Sequence=Displayed;  
CC Name=2;  
CC IsoId=Q86SR1-2; Sequence=VSP\_011209;  
CC Note=No experimental confirmation available;  
CC Name=3;  
CC IsoId=Q86SR1-3; Sequence=VSP\_011212, VSP\_011213;  
CC Note=No experimental confirmation available;  
CC Name=4;  
CC IsoId=Q86SR1-4; Sequence=VSP\_011207, VSP\_011208, VSP\_011214;  
CC Note=No experimental confirmation available;  
CC Name=5;  
CC IsoId=Q86SR1-5; Sequence=VSP\_011210, VSP\_011211;  
CC Note=No experimental confirmation available;  
CC -1- TISSUE SPECIFICITY: Widely expressed. Expressed at high level in  
CC small intestine, and at intermediate levels in stomach, pancreas,  
CC ovary, thyroid gland and spleen. Weakly expressed in other  
CC tissues.  
CC -1- DOMAIN: There are two conserved domains in the glycosyltransferase  
CC region: the N-terminal domain (domain A, also called Grl motif),  
CC which is probably involved in manganese coordination and substrate  
CC binding and the C-terminal domain (domain B, also called  
CC Gal/GlcNAc-T motif), which is probably involved in catalytic  
CC reaction and UDP-Gal binding (By similarity).  
CC -1- DOMAIN: The ricin B-type lectin domain binds to GalNAc and  
CC contributes to the glycopeptide specificity (By similarity).  
CC -1- SIMILARITY: Belongs to the glycosyltransferase family 2. GalNAc-T  
CC subfamily.  
CC -1- SIMILARITY: Contains 1 ricin B-type lectin domain.  
CC CAUTION: According to experiments made in rat, this enzyme is  
CC unable to transfer GalNAc onto serine or threonine residue on the  
CC protein receptor, but instead requires the prior addition of a  
CC GalNAc on a peptide before adding additional GalNAc moieties,  
CC thereby acting as a glycopeptide transferase.  
CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.ebi.ac.uk/announcements/>  
 CC or send an email to [license@ebi.ac.uk](mailto:license@ebi.ac.uk)).

CC -----

DR EMBL; AB078145; BACS6890.1; -  
 DR EMBL; AJ505950; CAD44532.1; -  
 DR EMBL; AL096739; CAB46378.1; -  
 DR EMBL; BC007224; AAH07224.2; -  
 DR EMBL; BC050333; AAH50333.1; -  
 DR EMBL; BC072450; AAH72450.1; -  
 DR EMBL; AK023782; BAB14676.1; ALT-INT.  
 DR EMBL; AK074132; BAB84958.1; -  
 DR PIR; T12552; T12552.  
 DR HSSP; P26514; IKNM.  
 DR Gene; HGNC:19873; GALNT10.  
 DR MIM; 608043; -  
 DR Interpro; IPR001173; Glyco\_trans\_2.  
 DR Interpro; IPR008997; RicinB-like.  
 DR Interpro; IPR000772; RicinB-lectin.  
 DR Pfam; PF00535; Glycos\_transf\_2; 1.  
 DR Pfam; PF00652; RicinB-lectin; 3.  
 DR SMART; SM00458; RICIN\_1.  
 DR PROSITE; PS50231; RICIN\_B-LECTIN; 1.  
 DR Alternative splicing; Calcium; Glycosyltransferase; Golgi stack;  
 KW Lectin; Manganese; Signal-anchor; Transferrase; Transmembrane.  
 FT DOMAIN 1 11  
 FT TRANSMEM 12 31  
 FT 31 Signal-anchor for type II membrane  
 FT 31 protein (Potential).  
 FT 31 Lumenal (Potential).  
 FT 31 Catalytic subdomain A.  
 FT 31 Ricin B-type lectin.  
 FT 31 By similarity.  
 FT 31 By similarity.  
 FT 31 N-linked (GlcNAc... ) (Potential).  
 FT 31 N-linked (GlcNAc... ) (Potential).  
 FT 31 N-linked (GlcNAc... ) (Potential).  
 FT 31 Missing (in isoform 4).  
 FT 31 /FTId=VSP\_011207.  
 FT 31 WELGYPGEIHWGEGEYISFK -> MLANRDELEAETS  
 FT 31 SSIFLAMO (in isoform 4).  
 FT 31 /FTId=VSP\_011208.  
 FT 31 Missing (in isoform 2).  
 FT 31 /FTId=VSP\_011209.  
 FT 31 EHLKPLADYVAL -> DLPTASTPSPVC (in  
 FT 31 isoform 5).  
 FT 31 /FTId=VSP\_011210.  
 FT 31 Missing (in isoform 5).  
 FT 31 /FTId=VSP\_011211.  
 FT 31 WMCGRMEDTCS -> SGLSRRLVLTAS (in  
 FT 31 isoform 3).  
 FT 31 /FTId=VSP\_011212.

Query Match 100.0%; Score 3278; DB 1; Length 603;  
 Best Local Similarity 100.0%; Pred. No. 3e-257;  
 Matches 603; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MRRERKLLQAVLALVALVLLPNVGLMALYRERQDPTPGSGAVALPAAGQSHSRQK 60  
 DB 1 MRRERKLLQAVLALVALVLLPNVGLMALYRERQDPTPGSGAVALPAAGQSHSRQK 60

QY 61 KTFPLGGGGLKQKMDKDEARRDQVNGGQGRPPYPTDARDOAVRENGFNIVYSDK 120  
 DB 61 KTFPLGGGGLKQKMDKDEARRDQVNGGQGRPPYPTDARDOAVRENGFNIVYSDK 120

QY 121 ISLNRSLPDIIRHPNCKSKRYLETLPNTSIIIPFNNEGSSLLRTVHSLVNSRPELVAEI 180  
 DB 121 ISLNRSLPDIIRHPNCKSKRYLETLPNTSIIIPFNNEGSSLLRTVHSLVNSRPELVAEI 180

QY 181 LVYDPSDRERHKKPLEDYVALFPVSRIILPTKREGILFRMLGASVATGDTITFLDHC 240  
 DB 181 LVYDPSDRERHKKPLEDYVALFPVSRIILPTKREGILFRMLGASVATGDTITFLDHC 240

QY 241 EAVNMLPPLLDRIARIRKTIIVCPMDIVIDHDPFRYETQAGDAMRGAFDMEYKRIPIP 300  
 DB 241 EAVNMLPPLLDRIARIRKTIIVCPMDIVIDHDPFRYETQAGDAMRGAFDMEYKRIPIP 300

QY 301 PELQKADPSDPFSPVYAGGLFAVDRKFWELGQYDGLIHWGEGEYISFKVMCGGRM 360  
 DB 301 PELQKADPSDPFSPVYAGGLFAVDRKFWELGQYDGLIHWGEGEYISFKVMCGGRM 360

QY 361 EDIPCSRVRGHIYKRYVYKYPAGVSLARNIKRAEAVMDYAEYIYQRRREYRLSGDV 420  
 DB 361 EDIPCSRVRGHIYKRYVYKYPAGVSLARNIKRAEAVMDYAEYIYQRRREYRLSGDV 420

QY 421 AVOKLSSLNCSFKFMTKIAMDLPKFYPPVPPAAGETIRNVGTGLCADTKHGALG 480  
 DB 421 AVOKLSSLNCSFKFMTKIAMDLPKFYPPVPPAAGETIRNVGTGLCADTKHGALG 480

QY 481 SPLREGCVRGRGAANNQVFTWRREDIRPGDPQHTKKCFDAISHTSPVTLVDCHS 540  
 DB 481 SPLREGCVRGRGAANNQVFTWRREDIRPGDPQHTKKCFDAISHTSPVTLVDCHS 540

QY 541 MKGNQMLKRYKDKTLVHPVSGSCMDGSESDHRIFMNTCNSSILQOQLFHTNSTVLEKF 600  
 DB 541 MKGNQMLKRYKDKTLVHPVSGSCMDGSESDHRIFMNTCNSSILQOQLFHTNSTVLEKF 600

QY 601 NRN 603  
 DB 601 NRN 603

RESULT 2  
 ID GLIO RAT STANDARD; PRT; 603 AA.  
 AC Q925R7;  
 DT 01-OCT-2004 (Rel. 45, Created)  
 DT 01-OCT-2004 (Rel. 45, Last sequence update)  
 DT 01-OCT-2004 (Rel. 45, Last annotation update)  
 DE Polypeptide N-acetylglucosaminyltransferase 10 (EC 2.4.1.41)  
 DE (Protein-UDP acetylglucosaminyltransferase 10) (UDP-  
 DE GalNAc:polypeptide N-acetylglucosaminyltransferase 10) (polypeptide  
 DE GalNAc transferase 10) (GalNAc-T10) (p-GANTase 10).  
 GN Name=Galnt10;  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A., ENZYME ACTIVITY, AND TISSUE SPECIFICITY.  
 RC TISSUE=Sublingual gland;  
 RX MEDLINE=21264503; Pubmed=11278534; DOI=10.1074/jbc.M009638200;  
 RA Ten Hagen K.G., Bedi G.S., Tetaert D., Kingsley P.D., Hagen F.K.,  
 RA Balya M.M., Beres T.M., Degand P., Tabak L.A.;  
 RT "Cloning and characterization of a ninth member of the UDP-  
 RT GalNAc:polypeptide N-acetylglucosaminyltransferase family,  
 RT pGANTase-T9.";  
 RT J. Biol. Chem. 276:17395-17404(2001).  
 RL J. Biol. Chem. 276:17395-17404(2001).  
 CC -1- FUNCTION: Catalyzes the initial reaction in O-linked  
 CC oligosaccharide biosynthesis, the transfer of an N-acetyl-D-  
 CC galactosamine residue to a serine or threonine residue on the  
 CC protein receptor. Has activity toward Muc5Ac and E2 peptide  
 CC substrates.  
 CC -1- CATALYTIC ACTIVITY: UDP-N-acetyl-D-galactosamine + polypeptide =  
 CC UDP + N-acetyl-D-galactosaminyl-polypeptide.  
 CC -1- COFACTOR: Manganese and calcium (By similarity).  
 CC -1- PATHWAY: Glycosylation.  
 CC -1- SUBCELLULAR LOCATION: Type II membrane protein. Golgi (By  
 CC similarity).  
 CC -1- TISSUE SPECIFICITY: Highly expressed in the sublingual gland,  
 CC testis, small intestine, colon and ovary. Expressed at

intermediate level in heart, brain, spleen, lung, stomach, cervix and uterus.

-1- DOMAIN: There are two conserved domains in the glycosyltransferase region: the N-terminal domain (domain A, also called GTI motif), which is probably involved in manganese coordination and substrate binding and the C-terminal domain (domain B, also called Gal/galnac-T motif), which is probably involved in catalytic reaction and UDP-Gal binding (By similarity).

-1- DOMAIN: The ricin B-type lectin domain binds to galnac and contributes to the glycopeptide specificity (By similarity).

-1- SIMILARITY: Belongs to the glycosyltransferase family 2. Galnac-T subfamily.

-1- SIMILARITY: Contains 1 ricin B-type lectin domain.

-1- CAUTION: According Ref.1, this enzyme is unable to transfer Galnac onto serine or threonine residue on the protein receptor, but instead requires the prior addition of a galnac on a peptide before adding additional galnac moieties, thereby acting as a glycopeptide transferase.

-1- CAUTION: Was originally (Ref.1) termed Galnac9/pp-Galnacase 9.

-----

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.iesb-sib.ch/announce/> or send an email to [license@iesb-sib.ch](mailto:license@iesb-sib.ch)).

-----

EMBL; AF241241; MAK54498.1; -.

DR HSBP; P26514; IKNL.

DR RGD; 69409; Galnac10.

DR InterPro; IPR001173; Glyco trans. 2.

DR InterPro; IPR008997; RicinB like.

DR InterPro; IPR000772; Ricin B lectin.

DR Pfam; PF00535; Glycosyltransf\_2; 1.

DR SMART; SM00652; Ricin B lectin; 3.

DR SMART; SM00458; Ricin; 1.

DR PROSITE; PS50231; RICIN B LECTIN; 1.

KM Calcium; Glycosyltransferase; Golgi stack; Lectin; Manganese;

KW Signal-anchor; Transferase; Transmembrane.

FT DOMAIN 1 11 Cytoplasmic (Potential).

FT TRANSMEM 12 31 Signal-anchor for type II membrane protein (Potential).

FT FT 32 603 Luminal (Potential).

FT DOMAIN 144 253 Catalytic subdomain A.

FT DOMAIN 311 373 Catalytic subdomain B.

FT DOMAIN 458 590 Ricin B-type lectin.

FT DISULFID 471 488 By similarity.

FT FT 523 538 By similarity.

FT DISULFID 563 578 By similarity.

FT CARBOHYD 124 124 N-linked (GLCNAC. . .) (Potential).

FT CARBOHYD 146 146 N-linked (GLCNAC. . .) (Potential).

FT CARBOHYD 593 593 N-linked (GLCNAC. . .) (Potential).

SQ SEQUENCE 603 AA; 69116 MW; 194EBA626A4BBF CRC64;

Query Match 96.3%; Score 3157; DB 1; Length 603;

Best Local Similarity 96.0%; Pred. No. 2e-247;

Matches 579; Conservative 7; Mismatches 17; Indels 0; Gaps 0;

QY 1 MRRREKRLQAVLALVLLPVGMLALYRERPDGTGGSGAANAAPAGQSGHSRQK 60

DB 1 MRRREKRLQAVLALVLLPVGMLALYRERPDGTGGSGAANAAPAGQSGHSRQK 60

QY 61 KTFPLSGQGLKDWHDKEAIRDAQRVNGEGQRPYPTMDAERVDQAARENGFNIVYSDK 120

DB 61 KTFPLSGQGLKDWHDKEAIRDAQRVNGEGQRPYPTMDAERVDQAARENGFNIVYSDK 120

QY 121 ISLNRSLPDIHRPNCNKRIVETLPTNTSIIIPFNHGGSSLLRTVHSLNRPPELVAEI 180

DB 121 ISLNRSLPDIHRPNCNKRIVETLPTNTSIIIPFNHGGSSLLRTVHSLNRPPELVAEI 180

QY 181 VLVDPSDRHLLKKPLEDYNALPPSVRIILRTKKRGLIRTMLGASVATGVITFLDSHC 240

DB 181 VLVDPSDRHLLKKPLEDYNALPPSVRIILRTKKRGLIRTMLGASVATGVITFLDSHC 240

DB 181 VLVDPSDRHLLKKPLEDYNALPPSVRIILRTKKRGLIRTMLGASVATGVITFLDSHC 240

QY 241 EAVNVMPLPLLDRIARARKTIVCPMIDVIDHDDPERYETQAGDAMRGAFDMEYKRIPI 300

DB 241 EAVNVMPLPLLDRIARARKTIVCPMIDVIDHDDPERYETQAGDAMRGAFDMEYKRIPI 300

QY 301 PELQKADPSDFESPVAAGGLFAVDRKFWELGQYDGLIWMGEQYISFKVMCGGRM 360

DB 301 PELQKADPSDFESPVAAGGLFAVDRKFWELGQYDGLIWMGEQYISFKVMCGGRM 360

QY 361 EDIPCSVGHITIKRYIVYKIPAGVSLARNIKRAVAEVMMDYAEITVYRRREYRLSAGDV 420

DB 361 EDIPCSVGHITIKRYIVYKIPAGVSLARNIKRAVAEVMMDYAEITVYRRREYRLSAGDV 420

QY 421 AVQKGLSSINCSFKKFMNTKIAMDLPKPYPPVEPPAAWGEIRNVTGTCADTKGALG 480

DB 421 AVQKGLSSINCSFKKFMNTKIAMDLPKPYPPVEPPAAWGEIRNVTGTCADTKGALG 480

QY 481 SPLRLGCGVGRGGAANNMQLFTWRREDIRPGDPQHTKKCPDAISHTSPVTLVDCGS 540

DB 481 SPLRLGCGVGRGGAANNMQLFTWRREDIRPGDPQHTKKCPDAISHTSPVTLVDCGS 540

QY 541 MKGNQMKRYKDKTLVHPVSGSCMDCSGDRIFMNTCNSSSLTQQLFHTNSTVLEKF 600

DB 541 MKGNQMKRYKDKTLVHPVSGSCMDCSGDRIFMNTCNSSSLTQQLFHTNSTVLENF 600

QY 601 NRN 603

DB 601 NRN 603

RESULT 3

BAD21405 PRELIMINARY; PRT; 634 AA.

AC BAD21405; 01-JUN-2004 (TrEMBLrel. 27, Created)

DT 01-JUN-2004 (TrEMBLrel. 27, Last sequence update)

DT 01-JUN-2004 (TrEMBLrel. 27, Last annotation update)

DE MFJ00205 protein (Fragment).

GN MFJ00205.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI\_TaxID=10090;

RN [1]

RP SEQUENCE FROM N. A.

RA Okazaki N., Kikuno R., Ohara R., Inamoto S., Koseki H., Hiraoka S., Saga Y., Kitamura H., Nakagawa T., Nagase T., Ohara O., Koga H.; "Prediction of the Coding Sequences of Mouse Homologues of FLJ Gene: The Complete Nucleotide Sequences of 110 Mouse FLJ-homologous cDNAs Identified by Screening of Terminal Sequences of cDNA Clones Randomly RT Amplified from Size-Fractionated Libraries."

RL Submitted (FEB-2004) to the EMBL/GenBank/DBJ databases.

DR EMBL; AKJ31155; BAD21405.1; -.

FT NON TER 1

SQ SEQUENCE 634 AA; 72315 MW; C79B82D4D0052C81 CRC64;

Query Match 95.7%; Score 3138; DB 2; Length 634;

Best Local Similarity 95.5%; Pred. No. 7.6e-246;

Matches 576; Conservative 9; Mismatches 18; Indels 0; Gaps 0;

QY 1 MRRREKRLQAVLALVLLPVGMLALYRERPDGTGGSGAANAAPAGQSGHSRQK 60

DB 32 MRRREKRLQAVLALVLLPVGMLALYRERPDGTGGSGAANAAPAGQSGHSRQK 91

QY 61 KTFPLSGQGLKDWHDKEAIRDAQRVNGEGQRPYPTMDAERVDQAARENGFNIVYSDK 120

DB 92 KTFPLSGQGLKDWHDKEAIRDAQRVNGEGQRPYPTMDAERVDQAARENGFNIVYSDK 151

QY 121 ISLNRSLPDIHRPNCNKRIVETLPTNTSIIIPFNHGGSSLLRTVHSLNRPPELVAEI 180

DB 152 ISLNRSLPDIHRPNCNKRIVETLPTNTSIIIPFNHGGSSLLRTVHSLNRPPELVAEI 211

QY 181 VLVDSEDRHLKKLEEDYALFPSPVRLTKRKGELIRTRMLGASVATGVTFLDSHC 240  
 Db 212 VLVDSDSDRHLKKLEEDYALFPSPVRLTKRKGELIRTRMLGASVATGVTFLDSHC 271  
 QY 241 EAVNNVLPILLDIRARRKRTIVCEMIDVIDHDDFRYETQAGDAMRGAEDWEMYKRIPIP 300  
 Db 272 EAVNNVLPILLDIRARRKRTIVCEMIDVIDHDDFRYETQAGDAMRGAEDWEMYKRIPIP 331  
 QY 301 PELQKADPSDPSPFPVWAGGLFANDRKFWELGSDYDGLHNGEJOYEISKVMCCGRM 360  
 Db 332 PELQKADPSDPSPFPVWAGGLFANDRKFWELGSDYDGLHNGEJOYEISKVMCCGRM 391  
 QY 361 EDIPCSRVGHIYRYKYVPYKYPAGVSLARNKRVAVNMDEFAEYIYORRPREYRLSAGDV 420  
 Db 392 EDIPCSRVGHIYRYKYVPYKYPAGVSLARNKRVAVNMDEFAEYIYORRPREYRLSAGDV 451  
 QY 421 AVQKRLSSLNCKSPKFMFKTAKIANDLPKYPVPVPPAAANGELIRNVTGLCADTKHAGLG 480  
 Db 452 VAQKRLSSLNCKSPKFMFKTAKIANDLPKYPVPVPPAAANGELIRNVTGLCADTKHAGLG 511  
 QY 481 SPLRECCVGRGGAANNNOVFTFTWRREDIRPDPQHTKFCPCDAISHTSPVTLYDCHS 540  
 Db 512 SPLRECCVGRGGAANNNOVFTFTWRREDIRPDPQHTKFCPCDAISHTSPVTLYDCHS 571  
 QY 541 MKGNQMLKRYRKDTLHYPVGSSCMDCSDRIIPMTCNBSLSLQOQLFHTNSTVLEKF 600  
 Db 572 MKGNQMLKRYRKDTLHYPVGSSCMDCSDRIIPMTCNBSLSLQOQLFHTNSTVLEKF 631  
 QY 601 NRN 603  
 Db 632 NRN 634

# RESULT 4 GL10 MOUSE STANDARD; PRT; 603 AA.

ID GL10 MOUSE STANDARD; PRT; 603 AA.  
 AC 06P957; 06KAO2; 06B208; 091Y66;  
 DT 01-OCT-2004 (Rel. 45, Created)  
 DT 01-OCT-2004 (Rel. 45, Last sequence update)  
 DE Polypeptide N-acetylgalactosaminyltransferase 10 (EC 2.4.1.41)  
 DE (Protein-UDP acetylgalactosaminyltransferase 10) (UDP-  
 GAlNAc:polypeptide N-acetylgalactosaminyltransferase 10) (polypeptide  
 DE GAlNAc transferase 10) (GAlNAc-T10) (pp-GANTase 10).  
 GN Name=GAlnt10;  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Okazaki N., Kikuno R., Ohara R., Inamoto S., Koseki H., Hiraoka S.,  
 RA Suga Y., Kitamura H., Nakagawa T., Nagase T., Ohara O., Koga H.;  
 RA "Prediction of the coding sequences of mouse homologues of Flj genes:  
 RT the complete nucleotide sequences of 110 mouse Flj-homologous cDNAs  
 RT identified by screening of terminal sequences of cDNA clones randomly  
 RT sampled from size-fractionated libraries.";  
 RL DNA Res. 11:167-180(2004).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6, and FVB/N. TISSUE=Brain, and Breast tumor;  
 RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;  
 RA Klausberg R.L., Collins E.A., Grouse L.H., Derge J.G.,  
 RA Klausberg R.D., Collins F.S., Wagner L., Shennan C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Hsieh F.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heien F.,  
 RA Diatchenko L., Marinska K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Brownstein M., Soares M.B., Bonaldo M.F., Casavant T.J., Scheetz T.E.,  
 RA Brownstein M.J., Udell T.B., Toshimiyuki S., Carninci P., Prange C.,  
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,  
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Wotley K.C., Hale S., Garcia A.M., Gay L.J., Bulik S.W.,  
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,

RA Fahney J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.W.,  
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smillie D.E.,  
 RA Schercher A., Schein J.B., Jones S.J.M., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length human  
 RT and mouse cDNA sequences.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 RN [3]  
 RP SEQUENCE OF 76-603 FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Colon;  
 RX MEDLINE=22354683; PubMed=12466851; DOI=10.1038/nature01266;  
 RA Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,  
 RA Nikaido I., Osato N., Saito R., Suzuki H., Yamanka I., Kiyosawa H.,  
 RA Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schonbach C., Gotoori T.,  
 RA Baldeceli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,  
 RA Schirral L.M., Kanapin A., Matsuda H., Batalov S., Beisel K.W.,  
 RA Blake J.A., Bradt D., Brusic V., Choithia C., Corbani L.E., Cousins S.,  
 RA Dalla E., Dragani T.A., Fletcher C.F., Forrest A., Frazer K.S.,  
 RA Gaasterland T., Gariboldi M., Gissi C., Godzik A., Gough J.,  
 RA Grimmond S., Gustincich S., Hirokawa N., Jackson I.D., Jarvis E.D.,  
 RA Kanai A., Kawaji H., Kawasawa Y., Kedzierski R.M., King B.L.,  
 RA Konagaya A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A.,  
 RA Maglott D.R., Maltais L., Marchionni L., McKenzie L., Miki H.,  
 RA Nagashima T., Numata K., Okido T., Pavan W.J., Perlea G., Pesole G.,  
 RA Petrovsky N., Pillai R., Pontius J.V., Qi D., Ramachandran S.,  
 RA Ravasi T., Reed J.C., Reed D.J., Reid J., Ring B.Z., Ringwald M.,  
 RA Sandelin A., Schneider C., Sempile C.A., Setou M., Shimada M.,  
 RA Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita M.,  
 RA Verardo R., Wagner L., Wahlestedt C., Wang Y., Watanabe Y., Wells C.,  
 RA Wilming L.G., Wynnshaw-Boris A., Yanagisawa M., Yang I., Yang L.,  
 RA Yuan Z., Zavolan M., Zhu Y., Zimmer A., Carninci P., Hayatsu N.,  
 RA Hirozane-Kishikawa T., Kono H., Nakamura M., Sakazume N., Sato K.,  
 RA Shiraki T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukuda S.,  
 RA Hara A., Hashizume M., Imotani K., Ishii Y., Itoh M., Kagawa I.,  
 RA Miyazaki A., Sakai K., Sasaki D., Shibata K., Shinagawa A.,  
 RA Yaunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J.,  
 RA Birney E., Hayashizaki Y.;  
 RT "Analysis of the mouse transcriptome based on functional annotation of  
 RT 60,770 full-length cDNAs.";  
 RL Nature 420:563-573(2002).  
 RN [4]  
 RP TISSUE SPECIFICITY  
 RX MEDLINE=21264503; PubMed=11278534; DOI=10.1074/jbc.M009638200;  
 RA Ten Hagen K.G., Bedi G.S., Tetaert D., Kingsley P.D., Hagen F.K.,  
 RA Baily M.M., Beres T.M., Degand P., Tabak L.A.;  
 RT "Cloning and characterization of a ninth member of the UDP-  
 RT GAlNAc:polypeptide N-acetylgalactosaminyltransferase family,  
 RT ppeantase-19.";  
 RL J. Biol. Chem. 276:17395-17404(2001).  
 CC -1- FUNCTION: Catalyzes the initial reaction in O-linked  
 CC oligosaccharide biosynthesis, the transfer of an N-acetyl-D-  
 CC galactosamine residue to a serine or threonine residue on the  
 CC protein receptor. Has activity toward MucAc and E2 peptide  
 CC substrates (By similarity).  
 CC -1- CATALYTIC ACTIVITY: UDP-N-acetyl-D-galactosamine + polypeptide =  
 CC UDP + N-acetyl-D-galactosaminyl-polypeptide.  
 CC -1- COFACTOR: Manganese and calcium (By similarity).  
 CC -1- PATHWAY: Glycosylation.  
 CC -1- SUBCELLULAR LOCATION: Type II membrane protein. Golgi (By  
 CC similarity).  
 CC -1- TISSUE SPECIFICITY: Expressed at higher level than GAlNT9. In the  
 CC developing hindbrain region of E14.5 embryos it accumulates in the  
 CC rapidly dividing, undifferentiated ventricular zone adjacent to  
 CC the pons. It also accumulates in the regions immediately rostral  
 CC and caudal to the dorsal rhombic lips differentiating into the  
 CC cerebellum. Not expressed in the developing choroid plexus.  
 CC -1- DOMAIN: There are two conserved domains in the glycosyltransferase  
 CC region: the N-terminal domain (domain A, also called G1 motif),  
 CC which is probably involved in manganese coordination and substrate  
 CC binding and the C-terminal domain (domain B, also called  
 CC Gal/GalNAc-T motif), which is probably involved in catalytic

```

CC reaction and UDP-Gal binding (By similarity).
CC -1- DOMAIN: The ricin B-type lectin domain binds to GAlNAc and
CC contributes to the glycopeptide specificity (By similarity).
CC -1- SIMILARITY: Belongs to the glycosyltransferase family 2. GAlNAc-T
CC subfamily.
CC -1- SIMILARITY: Contains 1 ricin B-type lectin domain.
CC CAUTION: According to experiments made in rat, this enzyme is
CC unable to transfer GAlNAc onto serine or threonine residue on the
CC protein receptor, but instead requires the prior addition of a
CC GAlNAc on a peptide before adding additional GAlNAc moieties,
CC thereby acting as a glycopeptide transferase.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.1sb-sib.ch/announce/
CC or send an email to license@1sb-sib.ch).
CC -----
DR EMBL; AK131155; BAD21405.1; ALT_INIT.
DR EMBL; BC016585; AAH16585.1; ALT_INIT.
DR EMBL; BC060617; AAH06017.1; -.
DR EMBL; AK033515; BAC28334.1; -.
DR HSSP; P26514; 1KRM.
DR MGD; MGI:1890480; Galnt10.
DR GO; GO:0004653; F:polypeptide N-acetylgalactosaminyltransferase. . .; IDA.
DR GO; GO:0006493; P:O-linked glycosylation; IDA.
DR InterPro; IPR001173; Glyco_trans_2.
DR InterPro; IPR008997; Ricin_like.
DR Pfam; PF00535; Glycos_transf_2; 1.
DR Pfam; PF00652; Ricin B_lectin; 2.
DR SMART; SM00458; RICIN_1.
DR PROSITE; PS50231; RICIN_B_LECTIN; 1.
KW Calicum; Glycosyltransferase; Golgi stack; lectin; Manganese;
KW Signal-anchor; transferase; Transmembrane.
FT DOMAIN 1 11 Cytoplasmic (Potential).
FT TRANSMEM 12 31 Signal-anchor for type II membrane
FT 32 603 protein (Potential).
FT DOMAIN 144 253 Lumenal (Potential).
FT DOMAIN 311 373 Catalytic subdomain A.
FT DOMAIN 458 590 Catalytic subdomain B.
FT DISULFID 471 488 Ricin B-type lectin.
FT DISULFID 523 538 By similarity.
FT DISULFID 563 578 By similarity.
FT CARBOHYD 124 124 N-linked (GlcNAc . . .) (Potential).
FT CARBOHYD 146 146 N-linked (GlcNAc . . .) (Potential).
FT CARBOHYD 593 593 N-linked (GlcNAc . . .) (Potential).
FT CONFLICT 233 233 V -> I (in Ref. 1).
SQ SEQUENCE 603 AA; 69116 MW; PF55FBA7E1DD7544 CRC64;

Query Match 95.7%; Score 3137; DB 1; Length 603;
Best Local Similarity 95.4%; Pred. No. 8.6e-246;
Matches 575; Conservative 10; Mismatches 18; Indels 0; Gaps 0;

```

```

QY 241 EAVNWMLPPLLDRIARRKRTIVCEMIDVIDHDDFRFYETOAGDAMRGAFDEMYKRIPIIP 300
DB 241 EAVNWMLPPLLDRIARRKRTIVCEMIDVIDHDDFRFYETOAGDAMRGAFDEMYKRIPIIP 300
QY 301 PELQKADPSPPPEBPWAGGLFAVDRKFWELGYPDGLITWGGEQVEISFKVMCCGRM 360
DB 301 PELQKADPSPPPEBPWAGGLFAVDRKFWELGYPDGLITWGGEQVEISFKVMCCGRM 360
QY 361 EDIPCSRVGHYIRCYVYKYPAGVSLARNLKRVAEVWMDVYAEYIYORRPEYRLISAGDV 420
DB 361 EDIPCSRVGHYIRCYVYKYPAGVSLARNLKRVAEVWMDVYAEYIYORRPEYRLISAGDV 420
QY 421 AVQKGLRSSLNCKSEFKFMFKTIAMDLPKYPVPEPPAAANGELRNVTGLCADTKRGALG 480
DB 421 AVQKGLRSSLNCKSEFKFMFKTIAMDLPKYPVPEPPAAANGELRNVTGLCADTKRGALG 480
QY 481 SPLRLBECVGRGGAANNMQVFTFRREDIRPDPQHTTKFCDAISHSPTLYCHS 540
DB 481 SPLRLBECVGRGGAANNMQVFTFRREDIRPDPQHTTKFCDAISHSPTLYCHS 540
QY 541 MKGNQLMKYRKDKTLVYPVSGSCMDSCSDHRIIPMTCNPSLSLQQLFEHTNSTVLEKF 600
DB 541 MKGNQLMKYRKDKTLVYPVSGSCMDSCSDHRIIPMTCNPSLSLQQLFEHTNSTVLEKF 600
QY 601 NRN 603
DB 601 NRN 603

RESULT 5
AAH60617 PRELIMINARY; PRT; 603 AA.
AC AAH60617;
DT 12-MAY-2004 (TReMBLrel. 27, Created)
DT 12-MAY-2004 (TReMBLrel. 27, Last sequence update)
DT 12-MAY-2004 (TReMBLrel. 27, Last annotation update)
DE UDP-N-acetyl-alpha-D-galactosamine:polypeptide
DE N-acetylgalactosaminyltransferase 10.
GN GALNT10.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N. A.
RC STRAIN=C57BL/6; TISSUE=Brain;
RX MEDLINE=22386257; PubMed=12477932;
RA Strauberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stadelton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Ueda T.B., Tohshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaby S.J.,
RA Boeak S.A., McGwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny K.C., Sodergren E.J., Lu X., Gibbs R.A.,
RA Faney J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Boulford G.G.,
RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Skelton J., Schmitt J., Myers R.M., Butterfield Y.S.,
RA Krywinski M.I., Skelton J., Small D.E., Schnerch A., Schein J.B.,
RA Jones S.J., Maira M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences."
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RP SEQUENCE FROM N. A.
RC STRAIN=C57BL/6; TISSUE=Brain;
RA Strauberg R.,
RL Submitted (OCT-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC060617; AAH60617.1; -.

```

KW	Transferrase.	603 AA:	69116 MW:	F855F8A7E1DD7544 CRC64:
SQL	SEQUENCE	603 AA:	69116 MW:	F855F8A7E1DD7544 CRC64:
	Query Match	95.7%;	Score 3137;	DB 2;
	Best Local Similarity	95.4%;	Pred. No. 8.6e-246;	
	Matches	575;	Conservative	10; Mismatches 18; Indels 0; Gaps 0
QY	1	MRRREKRLLOVALVALAALVLLPNTGMLATYREROPDGTGGSGAAVAAPAGCGSHSROK	60	
DB	1	MRRREKRLLOVALVALAALVLLPNTGMLATYREROPDGTGGSGAAVAAPAGCGSHSROK	60	
QY	61	KTFPLGAGDQCKDMHDKA1RRDARVGVGEGGRYPMTDARVDOAARENGFN1YVSXK	120	
DB	61	KTFPLGAGDQCKDMHDKA1RRDARVGVGEGGRYPMTDARVDOAARENGFN1YVSXK	120	
QY	121	ISLNSRLPD1RHPNCSKRYLETLTNTS111FFHNEGSSLLRTYHVSVLNRSPPELVAE1	180	
DB	121	ISLNSRLPD1RHPNCSKRYLETLTNTS111FFHNEGSSLLRTYHVSVLNRSPPELVAE1	180	
QY	181	VLVDDPFSREHLKPELIDYMALPESVR11RTYKRGEL1RTRM1GASVATGDVITFLDSHC	240	
DB	181	VLVDDPFSREHLKPELIDYMALPESVR11RTYKRGEL1RTRM1GASVATGDVITFLDSHC	240	
QY	241	EANVWMLPELLDRIARANKTIVCPMIDV1DHDDPFEYTOAGDAMGAPDMEYTRIP1P	300	
DB	241	EANVWMLPELLDRIARANKTIVCPMIDV1DHDDPFEYTOAGDAMGAPDMEYTRIP1P	300	
QY	301	PELOKADSDPEFSEVMAAGGLFPAVDRKFWELGCVDPGLE1WGEQYE1SFVMMCGGRM	360	
DB	301	PELOKADSDPEFSEVMAAGGLFPAVDRKFWELGCVDPGLE1WGEQYE1SFVMMCGGRM	360	
QY	361	EDIPCSRVGHIYRKTYPYKVPAGVSLARNLKXVAEVMDEVAEY1YORRPEYRLSAGDV	420	
DB	361	EDIPCSRVGHIYRKTYPYKVPAGVSLARNLKXVAEVMDEVAEY1YORRPEYRLSAGDV	420	
QY	421	AVQKKLRBSLNCSSKRWFEPTK1AMLPKPYPPVEEPPAAMGSIIRVVG1GLCDTTHGALG	480	
DB	421	AVQKKLRBSLNCSSKRWFEPTK1AMLPKPYPPVEEPPAAMGSIIRVVG1GLCDTTHGALG	480	
QY	481	SPLRLLEGCVRRGSGEAMANNMOVFETMRREDIRGDPPOHTKKCCPA1SHTSPTVL1YDCXS	540	
DB	481	SPLRLLEGCVRRGSGEAMANNMOVFETMRREDIRGDPPOHTKKCCPA1SHTSPTVL1YDCXS	540	
QY	541	MKGNO1WYRKDKTLYHPVSGSCMDSCSDSHRI1FNATCNPSL1TOO1LFEHTNS1VLEKF	600	
DB	541	MKGNO1WYRKDKTLYHPVSGSCMDSCSDSHRI1FNATCNPSL1TOO1LFEHTNS1VLEKF	600	
QY	601	NRN 603		
DB	601	NRN 603		
RESULT 6				
Q7Q0E9	PRELIMINARY;	PRT;	644 AA.	
AC	Q7Q0E9;			
DT	01-MAR-2004 (Tremblrel. 26, Created)			
DT	01-MAR-2004 (Tremblrel. 26, Last sequence update)			
DT	01-MAR-2004 (Tremblrel. 26, Last annotation update)			
DE	AGCP9480 (Fragment).			
GN	Name=agCG54007; ORFNames=ENSANG00000009226;			
OS	Anopheles gambiae str. PEST.			
OC	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;			
OC	Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae; Anopheles.			
NCBI	TaxID=180451;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=PEST;			
RL	Anopheles Genome Sequencing Consortium;			
CC	Submitted (MAR-2002) to the EMBL/Genbank/DBJ databases.			
CC	-1- CAUTION: The sequence shown here is derived from an			
CC	EMBL/Genbank/DBJ whole genome shotgun (WGS) entry which is			
CC	preliminary data.			

Query Match	42.8%	Score 1404;	DB 2;	Length 644;
Best Local Similarity	45.6%;	Pred. No. 5,6e-105;		
Matches 283;	Conservative 104;	Mismatches 195;	Indels 38;	Gaps 16
EMBL; AAA01008986; EAA00190.1; InterPro; IPR001173; Glyco_trans_2. InterPro; IPR00772; Ricin_B_lectin. Pfam; PF00535; Glycos_transf_2; 1. Pfam; PF00652; Ricin_B_lectin; 3. PROSITE; PS00231; R1CIN_B_LECTIN; 1. NON TER				
SEQUENCE 644 AA; 74214 MW; 087AB007BC228501 CRC64;				
1 MRKKERLLQAVLVYLAALVLLPNTGVLMALTYREROPDGTGGSGAAVAPAGQSGHSROK 60				
36 MRRNRVRLIKYLVFSGGALLFT-----TLIRSFSDNAKNTLTDGVFGPEAQMOKESHAR 89				
61 KTFPLGD-----GQVLKMDHKEAIRDAQVNGNEQGRPYPM--TDARVQAARENF 113				
90 EGSFNNNSKNVHQRIQIMWYNDLHEEAKSGSGEHSKAGQLDQSEHMKDLTKRGNF 149				
114 NIYVSDKSLNRSPLDIRHPNCSKRYLETLPNTSIIIPFNHEGSSILRTVSYLNRSP 173				
150 NAAVSDKSLNRSPLDIRHRCGRKKQYISELPTSVVVPFNEHSTILLRTASSVLRRP 209				
174 PELVAELVVDPSDREHLKKPLBEDYMA-LPPSVAILRTKKREGILRTRMIGASVATGDV 232				
210 PELIAEITLVDDCSFKEFLKQQLDEYVENMPKVAVRLPERSGLITARLAGAKIATADV 269				
223 ITPLDSHCEAVNMVLPILLDIRIARRKTIYCEPMIDVIDHDDPRYTOAGDARRGAPDWM 292				
270 LIFLDSSHTEAVNMVLPILLEPIAEDYRTCCVCEPFDVIDMDTFEYRAQ--DEGARGAFDWKF 328				
293 YYKRIIP-IPPELOKADPSDPFESPYPMAAGLFAVDKRMFMEIGYDPGLIHWGEQYEISF 351				
329 FYKRLPILPRDQ--NPTEPFESPMAAGLRAISAKFWMEIGYDEGLDINGGEQYEISF 386				
352 KVMNGGEMEDIPCSRVAHGIYRKVPYKVPAGVS-LAENLRVAEVMWDEYAEIYOR-R 409				
367 KIMQGGKMYDAPCSRVGHIYRGVAFPNRPKRDPLTRNYKRVAAEVMWDEYAEIYMRDR 446				
410 PEYRLSLGDVAVQKKLRSSLNCSFKKPFYKLIAMDLEKFPVPVPPPAAMKEIRNV-GT 468				
447 KKYETVTDGDISRQLAIREKQCKCFKFMFOVADLLEKYPILPPFPFANGAISOVANA 506				
469 GLCADTKGALGSPRLTEGVCVRGSEBAAMNMVQVTFPIWREDIRPGDPOHTKCF---CPD 525				
507 ALCVDTLHNGEKQITGLYSCABDKQPOPN--QFFQLSMHRDLR-----IKFELCMD 557				
526 AISHSP---VTLVDCHSMKGNQIMKYRKD-KTLVHPVSGSCMDCSBDRIFMNTCNPS 581				
558 -VSESVPAKILILYHCGGQGNQLMRYRPPQMLQKGQNNCLDMNPNNRBEVFPNCEPT 616				
562 SLTQGMLEFHTNSTVLEKFN 601				
617 NPROKRWGFTNATSLAQWN 636				
RESULT 7				
GL10_CABEL STANDARD; PRT; 622 AA.				
AC 045947;				
DT 01-OCT-2004 (Rel. 45, Created)				
DT 01-OCT-2004 (Rel. 45, Last sequence update)				
DT 01-OCT-2004 (Rel. 45, Last annotation update)				
DE Putative polypeptide N-acetylglucosaminyltransferase 10 (UDP-EC 2.4.1.41) (Protein-UDP acetylglucosaminyltransferase 10) (UDP-GalNAc:polypeptide N-acetylglucosaminyltransferase 10) (pp-GalNAc 10).				
GN Name=gly-10; ORFname=Y45FI0D.3;				
OS Caenorhabditis elegans.				
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae; Rhabditidae; Peloderinae; Caenorhabditis.				



RA Burtie K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahike C., Davenport L.B., Davies P.,  
 RA de Palos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.B., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Fextera S., Fleischmann W.,  
 RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Mei M.-H., Idegawa C.,  
 RA Jallali M., Kalush F., Karen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kamel B.E., Kodira C., Kraft C., Kravitz E., Kulp D., Lai Z.,  
 RA Laeko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,  
 RA Liu X., Matzel B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merklov G., Mithuna N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusekern D.R., Pacib J.M.,  
 RA Palazzolo M., Plutman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Slater E., Spadling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirbas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,  
 RT "The genome sequence of Drosophila melanogaster,"  
 RL Science 287:2185-2195(2000).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Berkley; TISSUE=Embryo;  
 RX MEDLINE=2242606; PubMed=12537569;  
 RA Stapleton M., Carlson J.W., Brokstein P., Yu C., Champagne M.,  
 RA George R.A., Garin H., Krommiller B., Pacib J.M., Park S., Wan K.H.,  
 RA Rubin G.M., Celniker S.E.,  
 RT "A Drosophila full-length cDNA resource,"  
 RL Genome Biol. 3:RESEARCH0080.1-RESEARCH0080.8(2002).  
 CC -1- FUNCTION: Glycopeptide transferase involved in O-linked  
 oligosaccharide biosynthesis, which catalyzes the transfer of an  
 N-acetyl-D-galactosamine residue to an already glycosylated  
 peptide. In contrast to other proteins of the family, it does not  
 act as a peptide transferase that transfers GalNAc onto serine or  
 threonine residue on the protein receptor, but instead requires  
 the prior addition of a GalNAc on a peptide before adding  
 additional GalNAc moieties. Some peptide transferase activity is  
 however not excluded, considering that its appropriate peptide  
 substrate may remain unidentified. Prefers the diglycosylated  
 Muc5AC-3/13 as substrate.  
 CC -1- COFACTOR: Manganese and calcium (By similarity).  
 CC -1- PATHWAY: Glycosylation.  
 CC -1- SUBCELLULAR LOCATION: Type II membrane protein. Golgi (By  
 similarity).  
 CC -1- TISSUE SPECIFICITY: In embryos, it is specifically expressed in  
 the salivary glands from stage 12, becoming stronger at stage 13.  
 CC Not expressed in other tissues. Further expressed during  
 oogenesis, in the somatically derived follicle cells that surround  
 the developing oocyte, which are involved in the maturation of the  
 oocyte and construction of the egg shell, as well as playing a  
 role in subsequent embryonic pattern formation.  
 CC -1- DEVELOPMENTAL STAGE: Expressed throughout embryonic, larval, pupal  
 and adult stages, with increasing levels during larval  
 development.  
 CC -1- DOMAIN: There are two conserved domains in the glycosyltransferase  
 region: the N-terminal domain (domain A, also called G1 motif),  
 which is probably involved in manganese coordination and substrate  
 binding and the C-terminal domain (domain B, also called  
 Gal/GalNAc-T motif), which is probably involved in catalytic  
 reaction and UDP-Gal binding (By similarity).  
 CC -1- DOMAIN: The ricin B-type lectin domain binds to GalNAc and  
 contributes to the glycopeptide specificity (By similarity).  
 CC -1- SIMILARITY: Belongs to the glycosyltransferase family 2. GalNAc-T  
 subfamily.  
 CC -1- SIMILARITY: Contains 1 ricin B-type lectin domain.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.ebi.ac.uk/announcements/>  
 CC or send an email to [license@ebi.ac.uk](mailto:license@ebi.ac.uk)).  
 CC -----  
 CC EMBL: AY268067; AA056703.1; -  
 DR EMBL: AE003476; AA47690.1; -  
 DR EMBL: AY061629; AA129177.1; -  
 DR FLYBASE: FBgn0035375; dgant6.  
 DR InterPro: IPR001173; Glyco\_trans\_2.  
 DR InterPro: IPR008997; RicinB\_1like.  
 DR InterPro: IPR000772; Ricin\_B\_lectin.  
 DR Pfam: PF00535; Glycos\_transf\_2; 1.  
 DR Pfam: PF00652; Ricin\_B\_lectin; 3.  
 DR SMART: SM00458; RICIN\_1.  
 DR PROSITE: PS50231; RICIN\_B\_LECTIN; 1.  
 KW Calcium; Glycosyltransferase; Golgi stack; Lectin; Manganese;  
 KW Signal-anchor; Transferase; Transmembrane.  
 FT DOMAIN 1 11  
 FT TRANSMEM 12 31  
 FT SIGNAL-ANCHOR 1 11  
 FT TRANSMEM 12 31  
 FT DOMAIN 32 666  
 FT DOMAIN 201 311  
 FT DOMAIN 367 429  
 FT DOMAIN 518 648  
 FT DISULFID 531 548  
 FT DISULFID 577 594  
 FT DISULFID 621 636  
 FT CARBOHYD 181 181  
 FT CARBOHYD 285 285  
 FT CARBOHYD 651 651  
 FT CARBOHYD 657 657  
 FT CONFLICT 95 96  
 FT CONFLICT 107 107  
 FT CONFLICT 499 499  
 FT SEQUENCE 666 AA; 76972 MW; CAECA6CE4860600C CRC64;  
 SQ  
 Query Match 41.5%; Score 1360; DB 1; Length 666;  
 Best Local Similarity 49.0%; Pred. No. 2.2e-101;  
 Matches 277; Conservative 76; Mismatches 160; Indels 52; Gaps 15;  
 QY DGQRLKQMDHKEAIRDQAGVNGEGQRPYMTDAERVDQAYR---ENGRNIYVSDKISTL 123  
 DB 121 DASVKQMDHYTMEKQAKVGVGEGKASTLDSEGRDLKEMSLENGFNALLSDSISV 180  
 QY 124 NRSLEPDIRPNCNSKRYLFTLPNTSTIIPPHNCGMSLTLTVHSVLRSPPELVATLV 193  
 DB 181 NRSVPDIRPDLCKRKEVAVLPVSVIIIFYNELSLVLRMSVSLNRSPELMKEIILV 240  
 QY 184 DDSDRHLLKKPLLEDVWA-LFSPVRIIRTKRKGRLRTMLGASVANGVITPLDSHCEA 242  
 DB 241 DDSDRHLLKKPLLEDVWA-LFSPVRIIRTKRKGRLRTMLGASVANGVITPLDSHCEA 300  
 QY 243 NVNMLPPLDIRIARNKRTICPMIDVIDHDFRYETQAGDAMGAPMWEYKRIPIPE 302  
 DB 301 NVNMLPPLDIRIARNKRTICPMIDVIDHDFRYETQAGDAMGAPMWEYKRIPIPE 359  
 QY 303 LQKADPEDPESPVPVAGLPAVDRKFMWELGGYDPLGELTNGSGOYRISFYVMCGGME 362  
 DB 360 DLK-HPADPEPKSPIMAGLPAISREFEWEIGYDEGLDINGGROYELSPKIMCGGMYD 418  
 QY 363 IPCSRRVGHIR-----KYVYKVPAGSLABNLRKRVAVMMDDYAEYIQRRPE-YRHL 416  
 DB 419 APCSRGHITRGPANHPSPRK--GQYLKKNTRKRVAVMMDDYAEYIQRRPE-YRHL 475  
 QY 417 AGDVAQKRLSSLNCKSPFMTKLAWDLPKPEVPVPPAAWAGETIRNVT-GLCADT- 474  
 DB 476 RGDLTEKAIKRTLNCKSPFMTKLAWDLPKPEVPVPPAAWAGETIRNVT-GLCADT- 535

QY	475	---	KRGALG	-----	SPRLREBCVAGRGSAAMNNQVFFETREDIRPDPDPOHTKKF	522				
Db	536	GRKKNNKGMATACADNITKTPOR	-----	---TQFWLSMKRDIR---	LRKKE	576				
QY	523	CFDA	-ISHTSPVTL	YDCHSMKNQMLMKY	-RKDKTL	YHPVSG	-SCMDCSES	DHRI	FMNTC	578
Db	577	CLDVQIWDANAPVWMLDCHSGCGNGYUWYDYDHNKQLKGTGSRGCTELL	PP	FGQEVVANK	636					
QY	579	NPSSLTQOMLEPHNTSYLEKRN	603							
Db	637	DTDNRFQOMNFGSFNKALTADNYSOD	661							
RESULT 9										
AAQ56703	PRELIMINARY;	PRT;	666	AA.						
AAQ56703	PRELIMINARY;	PRT;	666	AA.						
AC	AAQ56703;									
DT	02-MAR-2004	(TREMBlrel. 27, Created)								
DT	02-MAR-2004	(TREMBlrel. 27, Last sequence update)								
DT	02-MAR-2004	(TREMBlrel. 27, Last annotation update)								
DE	UDP-GalNAc:polypeptide N-acetylglactosaminyltransferase (EC 2.4.1.41).									
GN	PGANT6.									
OS	Drosophila melanogaster (Fruit fly).									
CC	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;									
CC	Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;									
CC	Ephydroidea; Drosophilidae; Drosophila.									
OX	NCBI_TaxID=7227;									
RN	[1]									
RP	SEQUENCE FROM N.A.									
RX	PubMed=1282974;									
RA	Ten Hagen K.G., Tran D.T., Gerken T.A., Stein D.S., Zhang Z.;									
RT	"Functional Characterization and Expression Analysis of Members of the									
RT	UDP-GalNAc:Polypeptide N-Acetylglactosaminyltransferase Family from									
RL	Drosophila melanogaster."									
RL	J. Biol. Chem. 278:35039-35048(2003).									
DR	EMBL: AY268067; AAQ56703.1; -									
KW	GLYCOSyltransferase; Transferase.									
SO	SEQUENCE 666 AA; 76958 MW; ID18362EB0DBD196 CRC64;									
Query Match 41.5%; Score 1360; DB 2; Length 666;										
Best Local Similarity 49.0%; Pred. No. 2.2e-101;										
Matches 277; Conservative 76; Mismatches 160; Indels 52; Gaps 15										
QY	67	DGQKLKDWHDKEAIRDQAVNGEGRPEMTDAERVDAYR--	ENGFNIVSDKISL	123						
Db	121	DASVKKDMDHYTFMEKDAKRVGLGEGKASTLDDEQRDLERKMSLENFALLSDSISV		180						
QY	124	NRSLLPDIHPNONSRYLETLPTNTSLIIFPHNEGSSLLRTQHSVYLNRSPPELVAIEIVY		183						
Db	181	NRSVVDIHHPICRKRKEVYAKLPVTSVLIIFVNEYSVLMRSVHSLNRSPPELMKEITLV		240						
QY	184	DDFSREHKKLEBDYMA-LFSPSVILRTKKEGGLRTMLDASVATGVITFLDSHCEA		242						
Db	241	DDHSREYELGKLELYTIAEHFKGVAVVRLPRRTGLIGAAAGARNNAITAVLLFLDSHVA		300						
QY	243	NVNMVLPILLDIRIARNRKTIVCPMIDVIDHDDFRFYEToGADAMRGAFDWMETTKRIPPE		302						
Db	301	NVNMVLPILLERFALNKRITVCPPIVIDITHTNHAYRAQ-DEGARGAFDMEFFYKRILPPE		359						
QY	303	LQKADSPDPEBSPVNAGGLFAVDRKMFWEIAGVDPGLETIWGSEOVYSIKVMMCGGRMD		362						
Db	360	DLK-HPADPKSPRIWAGGLFALISREFFWELGGVDEGLDIWGBQYELSKIKWCGGEMWD		418						
QY	363	IPCSRHGHIVR-----KVVPYKVPAGVSLARNLKVAAEVMDBEYAEIYQRPPE-VRHS		416						
Db	419	APCSRLIGHIYRPRNHQSPRK---GDYIAKHYYKVAEAYWDBEYKRYLLSHDGLYESVD		475						
QY	417	AGDVAVQKKLRSSLLNCKSPKFWMTKIAMDLPKFYRPVEBPAAWGEIRNVGT-GLCAQT-		474						
Db	476	PGDLTEQKRIFTKLNCKSPKFWMEVYAFDLMTYTPVDPDPSYAMGALQVGNQNLCLDTL		535						
QY	475	---KRGALG-----SPLREGCVRGRGEAAMNMVFTPTWREDIRPDPDPOHTKKF		522						

DB	536	GRKKHNMKGWYACADNIKT	TPGR-----TQFELSWKRLR-----LR	CKKE	576
QY	523	CFDA--ISHSPVTL	YDCHSKNGQWMTY-RKDTLHPVSG-SCMCS	SDHRI	FMNTC 578
DB	577	CLDVQIWMANP	PVIMWDCHSGGQWYVYDRHQLKHGTGR	RCLELLP	PSQEVVANKC 636
QY	579	NPSSLTOQWL	FEHTNSTVTL	KEFEN 603	
DB	637	DTNRFQDMNFG	FRKTLADNY	SOD 661	
RESULT 10					
Q7PZMS		PRELIMINARY;	PRT;	599	AA.
ID	Q7PZMS				
AC	Q7PZMS;				
DT	01-MAR-2004	(TREMBLrel. 26, Created)			
DT	01-MAR-2004	(TREMBLrel. 26, Last sequence update)			
DT	01-MAR-2004	(TREMBLrel. 26, Last annotation update)			
DE	AGCP9340	(Fragment).			
GN	Name=agcGS4449;	ORFNames=ENSANGC00000012813;			
OS	Anopheles gambiae	str. PEST.			
OC	Eukaryota;	Metazoa;	Arthropoda;	Hexapoda;	Insecta; Pezomyzeta;
OC	Neoptera;	Endopterygota;	Diptera;	Nematocera;	Culicoidae; Anopheles.
OX	NCBI_TaxID=180454;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=PEST;				
RA	Anopheles Genome Sequencing Consortium;				
RL	Submitted (MAR-2002) to the EMBL/Genbank/DBJ databases.				
CC	-1- CAUTION: The sequence shown here is derived from an				
CC	EMBL/Genbank/DBJ whole genome shotgun (WGS) entry which is				
CC	preliminary data.				
DR	EMBL; AAA801008986; EAA00339.1; -				
DR	InterPro; IPR001173; Glyco. trans. 2.				
DR	InterPro; IPR00772; Ricin B lectin.				
DR	Pfam; PF00535; Glycos. transf. 2; 1.				
DR	Pfam; PF00652; Ricin B lectin; 3.				
DR	PROSITE; PS50231; RICIN_B_LECTIN; 1.				
FT	NON TER				
FT	SEQUENCE 599 AA; 69224 MW; DA646C182B143028 CRC64;				
Query Match	40.9%;	Score 1341.5;	DB 2;	Length 599;	
Beet Local Similarity	45.8%;	Pred. No. 6e-100;			
Matches	257;	Conservative 104;	Mismatches 175;	Indels 25;	Gaps 13
DB	52	GOGSHRSKTKFFLDGQGLKDMWDKEAIRDADQVNGSGEGRPYMTDAERV--DQAY	108		
QY	52	GOGSHRSKTKFFLDGQGLKDMWDKEAIRDADQVNGSGEGRPYMTDAERV--DQAY	108		
DB	46	GEGFVAMPARNV---AGEKI-DMNYELIEBSKRTGFGHGRPKYKLSBQDIALMAKL	100		
QY	109	KENGNIYVSKISLNRSLPDIRHPNCSKRYLETLPNTSIIIPHNCSGLRTYASV	168		
DB	101	KENGSAVAVSDMIALNRSVDIRHPSCKMKYELKELPVSIIIFYNHWSALRTYVS	160		
QY	169	LNRSPELVARIIVLDPDSREBHLKKPLADYV--ALPSVAILRKTKREGILRTMLGAS	226		
DB	161	LNRSPELVARIIVLDPDSREBHLKKPLADYV--ALPSVAILRKTKREGILRTMLGAS	226		
QY	227	VATGQVITFLDSCHEANVMPLPLDLRIARAKRTIVCEMIDVIDIDHDDRYETQADAMRG	286		
DB	221	EARGGVLLVDSHTEVNTNMPLPLLEPRLAEDYRTVCDFIVIAIDTQVNSQ--DEGRKG	279		
QY	287	AFDWMYTKRIP-IPPELOKADPSDPESSPYMAGGLFVADRKMFYELGCGYDGLINWGE	345		
DB	280	AFDWMYTKRIP-IPPELOKADPSDPESSPYMAGGLFVADRKMFYELGCGYDGLINWGE	345		
QY	346	QYELSFKYMGGGRMEDIPCSRVGHIVKYVYKVPACVS-LARVLKRVAEVMDYAEV	404		
DB	348	QYELSFKYMGGGRMEDIPCSRVGHIVKYVYKVPACVS-LARVLKRVAEVMDYAEV	404		
QY	405	IYORPEYRHLASGVAVOQKLRSGLNCSKFEKMTKILAMLPKPEYVPEVPPAAMGEIR	464		
DB	398	IYERIPGQAKTDPDGLDSQRELRKELQCKPKFMFLVAVAPDLLVIRYPRPDQPFASGVQ	457		

```

QY 465 NVGT-GI-CADTKHGALGSPILREGCVRGRGAAMNNQVFTTWREDIRPDGPHQTKKFC 523
DB 458 SVANPRCLDLSLNQAKPEIGLYACAFNKTHP--ONNOFTLSYHRIIR---VRSNDKC 511
QY 524 FDAISHTSPVLYDCHSMKGNQLMKY-RKDKTYLHPVS--GSCMDCESDHRIFMNTCNP 580
DB 512 LDAKANDDELIVLFSCHSQGNQMRDYCESMIIHGKDHRCHMADLHTKLVRSCKD 571
QY 581 SSLTQOMLFHTNSTVLEKEN 601
DB 572 DKASQKWNMGYVNFVHLQNM 592

RESULT 11
Q7ODRO PRELIMINARY; PRT; 1003 AA.
AC Q7ODRO;
DT 01-MAR-2004 (Tremblrel. 26, Created)
DT 01-MAR-2004 (Tremblrel. 26, Last sequence update)
DE AccPI0665 (Fragment).
GN Name=sgCG47419; ORFNames=ENSANGC0000013497;
OS Anopheles gambiae str. PEST.
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae; Anopheles.
OX NCBI_TaxID=180454;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PEST;
RA Anopheles Genome Sequencing Consortium;
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR EMBL; AAAB01008849; EAA07231.1; -.
DR InterPro: IPR001173; Glyco_trans_2.
DR InterPro: IPR000772; Ricin_B_lectin.
DR Pfam: PF00535; Glycos_transf_2; 2.
DR Pfam: PF00652; Ricin_B_lectin; 4.
DR PROSITE; PS50231; RICIN_B_LECTIN; 2.
FT NON TER
FT 1
SQ SEQUENCE 1003 AA; 115923 MW; 753EA50F567AAB13 CRC64;

Query Match 38.0%; Score 1247; DB 2; Length 1003;
Best Local Similarity 47.6%; Pred. No. 5.9e-92;
Matches 258; Conservative 77; Mismatches 179; Indels 28; Gaps 14;

QY 73 DMHDKAIRDQAVNGEGGRPYMTDARDV---QAYRANGFNIVYSKISLNSLP 128
DB 27 DYNHYEDIQNDLNKVGEGQKPYTISPEEXTSELRELYKGFNLSDKISINSIA 86
QY 129 DIHHPCNSKRYLETTPNTSIIIPHNHGNSSLLRTVSHVNLNRPPELVAVIVDFSD 188
DB 87 DLNHPCKLKSYSNHLPIASVVVFYEHNSTLRTIYSVLNRPPLHKEIIIVDGGST 146
QY 189 REHKKPLEDYMAL-PPSVRLRTKKKGLIRTEMLGASVATGVITFLDSHCANVNL 247
DB 147 KEFLHNKLEIDYVKNLPRKVLVROPERTGLIKALAGAKIASGVLLFLDSHTAGVNL 206
QY 248 PLLLDIRIARKRTIVCMIVDIDHDDFRYEQADARGAADVMYTKRIPPELOKAD 307
DB 207 PLLLEPIAENPKTCVCPILVIDQTDVHPQ--DEGRGFLDWFHYKRVVKNK-DRIS 264
QY 308 PSDPFESPVAAGLFAVDRKFMFELAGYDPLIWMGEQYEISFKVMMCGRMEDICSR 367
DB 265 PTEPPEPVAAGLFAIGADPFWELGSDYDELDIWMGEQYEISFKIQCCGRMLDACS 324
QY 368 VGHYRKTVYKVPAGVS-LARNIKRVAEVMDEYAEYIYORDE-YRHSAGVAVQKK 425
DB 325 FGHIIRYVSPFPNRSKYDFTIRNHNKRAEIMDEYKYIYIDRDERYAKTDAGDMSKKT 384
QY 426 LRSLINCKSPKMFMTKIAMLDPKYPVPEPPAAMGRIWNG-TGLCADTKHGALGSPILR 484

```

```

DB 385 IRETLMKPFRKMFLOEFAPEILIEYPPVEPEPVYASGSLQSVADSLCIDTMQRGRGEIG 444
QY 485 LEEGCVRGGAAMNNQVFTTWREDIRPDGPHQTKKFC--CPDAISHT--SPVLYDCH 539
DB 445 LYPSCNSLIEPT-NHNGYFVYSHWRDI-----QR-KYGECCFDPVQSPKSPVLIPTCH 496
QY 540 SMKGNQIMTKRKDKTYLHPVSGSCMD-----CSESDHIFMNTCNPSSLTQOMLFHTNST 595
DB 497 MHQGNQFPQY-DHKTQQIKRNGVCDSDPAKVECHRINYSKULPQTSIIVPFDEHWST 555
QY 596 VL 597
DB 556 LL 557

RESULT 12
GLT4_DROME STANDARD; PRT; 659 AA.
AC Q81A42; Q81Q11;
DT 01-OCT-2004 (Rel. 45, Created)
DT 01-OCT-2004 (Rel. 45, Last sequence update)
DE 01-OCT-2004 (Rel. 45, Last annotation update)
DT 01-OCT-2004 (Rel. 45, Last annotation update)
DE N-acetylglucosaminyltransferase 4 (EC 2.4.1.-) (Protein-UDP
DE acetylglucosaminyltransferase 4) (UDP-GalNAc:polypeptide N-
DE acetylglucosaminyltransferase 4) (pp-GalNAc 4).
GN Name=Pgant4; ORFNames=CG31956;
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A., ENZYME ACTIVITY, TISSUE SPECIFICITY, AND
RP DEVELOPMENTAL STAGE.
RC STRAIN=Canton-S; TISSUE=Embryo;
RX MEDLINE=2884110; PubMed=12829714; DOI=10.1074/jbc.M3036200;
RA Ten Hagen K.G., Tran D.T., Gerken T.A., Stein D.S., Zhang Z.;
RT "Functional characterization and expression analysis of members of the
RT UDP-GalNAc:polypeptide N-acetylglucosaminyltransferase family from
RT Drosophila melanogaster."
RL J. Biol. Chem. 278:35039-35048 (2003).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkely;
RX MEDLINE=20196006; PubMed=10731132; DOI=10.1126/science.287.5461.2185;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hopkins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abell J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballow R.M., Basu A., Baxendale J., Bayraktarglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Botkova D., Botchan M.R., Bouck J., Brockeisen P., Brothier P.,
RA Butcher K.C., Busam D.A., Butler H., Cadietu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Fowler C., Gabriellian A.B., Gary N.S., Gelbart W.M., Glasser K.,
RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jaitai M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Laeko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Meuklov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nushekn D.R., Pacלב J.M.,
RA Palazzo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,

```



QY 506 TWREDIRPGDPOHTKKE---CFDAISHTSP-----VTLYDCHSMKGNQLMKY-RKDKTLTYH 557  
 DB 557 SIFRDLR-----MKGFDVSLDV--HEGPNATVMMWSCHSGGNGFWYDROTQLRVH 608  
 QY 558 PVSGS-CMD--CSESDHRIFMNTGNPSLSLTOQWLFHTNSTVLEKEF 600  
 DB 609 GENNKRCLEGFVENGIKAVVANSCEGNDQRMWFGFVNHTMLDTF 654

RESULT 13  
 AA056701 PRELIMINARY; PRT; 659 AA.  
 ID AA056701  
 AC AA056701:  
 DT 02-MAR-2004 (Tremblrel. 27, Created)  
 DT 02-MAR-2004 (Tremblrel. 27, Last sequence update)  
 DT 02-MAR-2004 (Tremblrel. 27, Last annotation update)  
 DE UDP-GalNac:polypeptide N-acetylglactosaminyltransferase (EC 2.4.1.41).  
 GN PGANT9.  
 OS Drosophila melanogaster (fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OC NCBI\_TaxId=7227;  
 [1]  
 RP SEQUENCE FROM N.A.  
 RX PubMed=1829714;  
 RA Ten Hagen K.G., Tran D.T., Gerken T.A., Stein D.S., Zhang Z.;  
 RT "Functional Characterization and Expression Analysis of Members of the  
 RT UDP-GalNac:polypeptide N-acetylglactosaminyltransferase Family from  
 RT Drosophila melanogaster";  
 RL J. Biol. Chem. 278:35039-35048 (2003).  
 DR EMBL: AY268065; AA056701.1; -  
 KW Glycosyltransferase; Transferase.  
 SQ SEQUENCE 659 AA; 75805 MW; 0182D42A8BA1DD2 CRC64;

Query Match 36.2%; Score 1186; DB 2; Length 659;  
 Best Local Similarity 41.2%; Pred. No. 2.9e-87;  
 Matches 266; Conservative 106; Mismatches 216; Indels 58; Gaps 21;

QY 2 RRKEKRLQAVLALVLLPNVGLMALYR-----EROPDG---TPGSGAAYA 48  
 DB 20 KRYVKKRLKRVLLVIVVSVLTIVVERMKMAALETRQDLPNDPITVPRANIH 79  
 QY 49 PA--AGQSHSRQKTFPLDGGQTL-----KDWKKAIRRDQAVNGEQGR 94  
 DB 80 PTKKAPPPQDNRSVVDIPRSDLQGFRLPEPKGRKMDWDYAAEADRRKSGFGEHV 139  
 QY 95 PYRM--TDARVD-QAVRENGFNIVYDSTLSRSLDIPRNCNSKRYLETLPNTSITI 151  
 DB 140 AVKIEPDERQOLEKEHEKEMNGFNGLISDTSVNSVSDLRLEACKTKRYLAKLPNISVIF 199  
 QY 152 PFNHEGSSILRTVSVLNSRSPPELVAVIYLVDDPSREHLKKPLBRYMAL-FPS-VRIL 209  
 DB 200 IFNNEHNTILKRSYVYNTTPPELAKQIVLVDDGSEMDVLKQGLDIYVQHPHVLTV 259  
 QY 210 RTTKREGLITRMLGASVATGDTITFLDSHCENANVMPLLDRIARNRRTIVCPMDIVI 269  
 DB 260 RNEPQGLIGARLAGAKVAVGVWVFPDISHLEVYVMPLPLIEPIALINPKISTCPMDVTI 319  
 QY 270 DHDDFERETQAGAMRAFPWEMTYKRIPIPELQKADPSDPPSPMAGGLPVDKMF 329  
 DB 320 SHEDFSYFSGNKGARGGFWKMLYKQIPVLPF-DALDKMPYKSPVMGGLFINTDF 378  
 QY 330 WEIGGYDPLGEIWCGEQYISIFKVMCGGRMEDIPGSRVGIYV-KVVPKVPAGVS-LA 387  
 DB 379 WDLGGVDQDLIDINGEYELSFKIMCGGMLDVPGRVAHTFGPKPKPGKPNPGRNPA 438  
 QY 388 RNLKRAVAVMWDEAYEYIYORRPE-YRHLGADVAVOKLRSLSLNCKSFKFMFKTIAMD 446  
 DB 439 KHKRAVAVVMDEYKQVYVYKRDPTKYTNLDAGDLTRQGVRELRKCKSGFHMFTVEVAPDF 498

QY 447 KTFYPPVPPAAAMGEIRNVGTGL-CADTKGALGSPLEGGYRGGBAAMNMOYETF 505  
 DB 499 LVKFPVPPSPSYAAGIIQVANVPYCLDNNKGS TEAVGHSFQADNRTHQPN--QWEL 556  
 QY 506 TWREDIRPGDPOHTKKE---CFDAISHTSP-----VTLYDCHSMKGNQLMKY-RKDKTLTYH 557  
 DB 557 SIFRDLR-----MKGFDVSLDV--HEGPNATVMMWSCHSGGNGFWYDROTQLRVH 608  
 QY 558 PVSGS-CMD--CSESDHRIFMNTGNPSLSLTOQWLFHTNSTVLEKEF 600  
 DB 609 GENNKRCLEGFVENGIKAVVANSCEGNDQRMWFGFVNHTMLDTF 654

RESULT 14  
 ID GT9\_DROME STANDARD; PRT; 650 AA.  
 AC OSMRC9; Q9V770;  
 DT 01-OCT-2004 (Rel. 45, Created)  
 DT 01-OCT-2004 (Rel. 45, Last sequence update)  
 DT 01-OCT-2004 (Rel. 45, Last annotation update)  
 DE Putative polypeptide N-acetylglactosaminyltransferase 9 (EC 2.4.1.41)  
 DE (Protein-UDP acetylglactosaminyltransferase 9) (UDP-  
 DE GalNac:polypeptide N-acetylglactosaminyltransferase 9) (p-  
 DE 9).  
 GN Name=pgant9; ORFNames=CG30463;  
 OS Drosophila melanogaster (fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OC NCBI\_TaxId=7227;  
 [1]  
 RP SEQUENCE FROM N.A.  
 RX STRAIN=Berkley;  
 RX MDL=20196006; PubMed=10731132; DOI=10.1126/science.287.5461.2185;  
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Ahril J.F., Agbayani A., An H.-D., Andrews-Pfannkoch C., Baldwin D.,  
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Bernan B.P., Bhandari D., Bolshakov S.,  
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotler P.,  
 RA Burtis K.C., Busam D.A., Butler H., Cadiou E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson R., Dou P.L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
 RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Idegam C.,  
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kemison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lasko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J.M., Mostreli A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleby J.M.,  
 RA Palazolo M., Pittman G.S., Pan S., Pollard S., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skyski M.P., Smith T.,  
 RA Slier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Sytkies A., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wasmann D.A., Weinstock G.M., Weisenbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye Y., Yen R.-F., Zavari J.S., Zhan W., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers B.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of Drosophila melanogaster";  
 RL Science 287:2185-2195(2000).  
 RN [2]

RE	REVIEWS:	
RE	MEDLINE=22426069; PubMed=12537572;	
RE	Mitra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,	
RE	Hradecky P., Huang Y., Kaminker J.S., Milburn G.H., Prochuk S.E.,	
RE	Smith C.D., Tapp J.L., Whitfield E.J., Bayraktaroglu L., Bernan B.P.,	
RE	Bettencourt B.R., Colinker S.E., de Grey A.D.N.J., Drysdale R.A.,	
RE	Harris N.L., Richter C., Russo S., Schroder A.J., Shu S.Q.,	
RE	Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,	
RE	Lewis S.E.;	
RE	"Annotation of the Drosophila melanogaster euchromatic genome: a	
RE	systematic review."	
RE	Genome Biol. 3:RESEARCH0083.1-RESEARCH0083.22(2002).	
RE	[3]	
RE	SEQUENCE FROM N.A.	
RE	STRAIN=Berkeley; TISSUE=Embryo;	
RE	RE MEDLINE=22426066; PubMed=12537569;	
RE	Stapleton M., Carlson J.W., Brokerstein P., Yu C., Champagne M.,	
RE	George R.A., Guatin H., Krommiller B., Pacled J.M., Park S., Wan K.H.,	
RE	Rubin G.M., Colinker S.B.;	
RE	"A Drosophila full-length cDNA resource."	
RE	Genome Biol. 3:RESEARCH0080.1-RESEARCH0080.8(2002).	
RE	-1- FUNCTION: May catalyze the initial reaction in O-linked	
RE	oligosaccharide biosynthesis, the transfer of an N-acetyl-D-	
RE	galactosamine residue to a serine or threonine residue on the	
RE	protein receptor (By similarity).	
RE	-1- CATALYTIC ACTIVITY: UDP-N-acetyl-D-galactosamine + polypeptide =	
RE	UDP + N-acetyl-D-galactosaminyl-polypeptide.	
RE	-1- COFACTOR: Manganese and calcium (By similarity).	
RE	-1- PATHWAY: Glycosylation.	
RE	-1- SUBCELLULAR LOCATION: Type II membrane protein. Golgi (By	
RE	similarity).	
RE	-1- DOMAIN: There are two conserved domains in the glycosyltransferase	
RE	region: the N-terminal domain (domain A, also called GTI motif),	
RE	which is probably involved in manganese coordination and substrate	
RE	binding and the C-terminal domain (domain B, also called	
RE	Gal/GalNAc-T motif), which is probably involved in catalytic	
RE	reaction and UDP-Gal binding (By similarity).	
RE	-1- DOMAIN: The ricin B-type lectin domain binds to GalNAc and	
RE	contributes to the glycopeptide specificity (By similarity).	
RE	-1- SIMILARITY: Belongs to the glycosyltransferase family 2. GalNAc-T	
RE	subfamily.	
RE	-1- SIMILARITY: Contains 1 ricin B-type lectin domain.	
RE	-----	
RE	This SWISS-PROT entry is copyright. It is produced through a collaboration	
RE	between the Swiss Institute of Bioinformatics and the EMBL outstation -	
RE	the European Bioinformatics Institute. There are no restrictions on its	
RE	use by non-profit institutions as long as its content is in no way	
RE	modified and this statement is not removed. Usage by and for commercial	
RE	entities requires a license agreement (see <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>	
RE	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).	
RE	-----	
RE	EMBL: AE003806; AAF57964.2; -	
RE	EMBL: AY121661; AAM51988.1; -	
RE	HSSP: P26514; IKM.	
RE	FlyBase: FBgn0050463; CG30463.	
RE	InterPro: IPR001173; Glyco. trans. 2.	
RE	InterPro: IPR009061; Putative DNA-bind.	
RE	InterPro: IPR008997; Ricin-like	
RE	InterPro: IPR00772; Ricin_B_lectin.	
RE	PIfam: PF00535; GlycoE_transf_2; 1.	
RE	PIfam: PF00652; Ricin_B_lectin; 3.	
RE	SMART: SM00458; RICIN; 1.	
RE	PROSITE: PS50231; RICIN_B_LECTIN; 1.	
RE	CaLectin, Glycosyltransferase, Golgi stack; Lectin; Manganese;	
RE	Signal-anchor; Transferase; Transmembrane.	
RE	Cytoplasmic (Potential).	
RE	Signal-anchor for type II membrane	
RE	protein (Potential).	
RE	luminal (Potential).	
RE	Catalytic subdomain A.	
RE	Catalytic subdomain B.	
RE	Ricin B-type lectin.	
RE	Gly-rich.	

PT	DISTLFTD	535	554		By similarity.
FT	DISTLFTD	577	590		By similarity.
FT	DISTLFTD	616	631		By similarity.
FT	CAROHYD	321	321		N-linked (GlcNAc. . .) (Potential).
FT	CAROHYD	373	373		N-linked (GlcNAc. . .) (Potential).
FT	CONFLICT	145	145		L -> D (in Ref. 3).
FT	CONFLICT	454	454		N -> P (in Ref. 3).
FT	CONFLICT	472	472		Y -> C (in Ref. 3).
SQ	SEQUENCE	650 AA;	73192 MM;	AC847736ADIC07CA CRC64;	
<hr/>					
Query Match		35.8%;	Score 1172.5;	DB 1;	Length 650;
Best Local Similarity		40.2%;	Pred. No.78-86;		
Matches		272;	Conservative	91;	Mismatches 201; Indels 113; Gaps 22.
<hr/>					
Oy	2 RKRKRLQAVALVLAALVLLPNVLGYEROPDGTGGSGAAVAAPAGQS-----	55			
Db	6 RRRSTTIVKLVAFLAIFWCIAPLVYTDDTRRRAAQEGASGASGAPGVGGAGGLDP	65			
Oy	56 -----HSROKKTFPL-----GDQQLKXMDH-----	76			
Db	66 IALALRNPAGEDPFGINGNVIIGGGQQ--KQAHDEADIPRTVGKHAKDLQAEEMRKKAEQ	123			
Oy	77 -KEAIRBPAGV-----GNGBQGSPY-----PMTD--AEAYDAQAREGNPIVYSKD	120			
Db	124 PKKPEQESQKYVIDEPPANFEENPGSLGPRVLPRKMSSEMKKAVDDCGTTKRAFQVQISDL	183			
Oy	121 ISLRSLPDIRHPNCSNK-RYLETLTPNTSIIIPHNESGSILRTVHSVLSRSPEELVAE	179			
Db	184 ISVHTLPDPDAMCKDEARLTMLPKTDVLIICHNEAMTVLLRTVHSVLDPSPEHLIGK	243			
Oy	180 IVLVDSFDRHLKKPLEEDYNALFPSVRILATFKREGILRTFMIGASYATGDVITPLDSH	239			
Db	244 IILVDVSDMDELTKOLEDYFAAPAPKVQIIRGOQRKEGLIRARIIGANNAPKSPVLYLDSH	303			
Oy	240 CEANVMVPELLDIRARARKTIQCMIDVIDHDPRYE-TQAGDMGA FPMEMYVKRI P	298			
Db	304 CECTEGMLEPLLDIRARNSTTVCCVIDVISDETLETLEYRDSGGVNGGFPMNLQFSWHP	363			
Oy	299 IPPELOKADPS--DFESPVAAGGLFAVDRKMFWELGSDPGLIEWGEQYEISFKVMWC	356			
Db	364 VPERRKKNHTABEVYSPMAAGLFSTRDRFPRLGTGYDGDGPIMWGSENLELFKTMWC	423			
Oy	357 GGREDIPCSRNGHIYKRYVPYKVPAGVS-LARNLKRAYAWMBDAEYIYQY---RPE	411			
Db	424 GGTLEIVPCSHVGHFRKSRPYKRSQNVNKKJNSVRLAEVWMBSQYYHYHRIGNDKGD	483			
Oy	412 YRHLSAGVNAVOKLARSSLNCSLKSPKMFMTKIAMLDPKRYPVDERP-AAANGELRVNOTG	469			
Db	484 W-----GVSSORRKLRLNDLKCFKSFMYLDNI-----YPRLFIPGDSVAAGELRINLDYG	531			
Oy	470 --LCADTYHG-----ALGPSRLBEGCYRGGEAAMNMNQVFTTWBEDIRPGPOHTKC	521			
Db	532 GRCLIDAPAGKKQKQKAVGT-----YPCIRGOGNQTW-----MISKAGEIRRDSD-----	576			
Oy	522 FCFDAISHTSPVTLYDCISMKGNDLMKYRKD-KTLVHPVSGSCNDCESDHRIFMANTCP	580			
Db	577 -CLDYAG--KDVTLFGCHGCKGNGFWTRYRENTKQLHGTSGKCIASISBKDKLMEBCSA	633			
Oy	581 SSLVYQOMLFPHNTSNVL	597			
Db	634 SLSROQMTELENYDSSKL	650			
<hr/>					
RESULT 15					
ID	Q70815	PRELIMINARY,	PRT,	645 AA.	
AC	Q70815;				
DT	01-MAR-2004 (TrEMBLrel. 26, Created)				
DT	01-MAR-2004 (TrEMBLrel. 26, Last sequence update)				
DT	01-MAR-2004 (TrEMBLrel. 26, Last annotation update)				
DE	EbiP5044 (fragment).				
OS	Name=ebiG5044; ORFNames=ENSANG0000003900;				
OS	Anopheles gambiae str. PE87.				

OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae; Anopheles.  
 OC NCBI\_TaxID=180454;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=PEST;  
 RA Anopheles Genome Sequencing Consortium;  
 RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.  
 CC -!- CAUTION: The sequence shown here is derived from an  
 EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is  
 preliminary data.  
 DR EMBL; AAAB01008944; EAA10180.1; -;  
 DR InterPro: IPR001173; Glyco\_transf\_2;  
 DR InterPro: IPR000772; Ricin\_B\_lectin.  
 DR Pfam; PF00535; Glycos\_transf\_2; 1.  
 DR Pfam; PF00652; Ricin\_B\_lectin; 3.  
 DR PROSITE; PSS0231; RICIN\_B\_LECTIN; 1.  
 DR NON\_TER 645 645  
 FT SEQUENCE 645 AA; 73459 MW; AADPDB6CE16B51C0 CRC64;

Query Match 35.4%; Score 1159; DB 2; Length 645;  
 Best Local Similarity 44.3%; Pred. No. 4,4e-85;  
 Matches 251; Conservative 79; Mismatches 193; Indels 44; Gaps 19;

QY 38 GTGGGGAATAAPAAAGGSHSRQKTFPLGSGQKXKMDHDEKATIRDAQVGN-GEGRP- 95  
 DB 101 GDEPGYGG--GGGRADSSMPRTY--RPQELKKRQAFTV--KENYGRPEWGPV 151  
 QY 96 -YPMTDARVDAQVRENGFNIVYSDKISLNRSLPDIRHPNCNSKRYLETPTNTSIIIPFH 154  
 DB 152 KIRANQGEMLKEKKEKQNFILASDMITWNRSLTDVHHQCKKHAYAKLPTTSIVAFH 211  
 QY 155 NEGSSSLRTVHSVYLNPSPELVAEIVLVDPSDRHKKPLEDYALFP-SVRIITPKK 213  
 DB 212 NEAMSTLRTIWSYVINSPPLEKEIILVDASEREHSGQLBEYVTLTPVTFVLTGK 271  
 QY 214 REGIIRRMIGASATGDTITFLDSHCEAVNMLPILLDIRARRKTIYCPMTDIVDHD 273  
 DB 272 RSGIIRRLIGAKKVKQGVITFLDACECTEGMEPLIARIVLDRKTVCCPIIDVDET 331  
 QY 274 FRYETQAGDAMRGAFDEMYTKRIPRP-ELQKA--DPSDPFESPVAAGLFAVDRKMF 330  
 DB 332 FEVYLT-ASDQTWGFNMKLPFRWRYVARERQNRNHTAPLRTPMAAGLFSIDRYFY 390  
 QY 331 ELGQYDPLGIEWGEQYEISPFKVMCGRMEDIPCSRVGHIYRKVYKYPAGVS--LAR 388  
 DB 391 EISYDEGMDIWGENLEMSFRVVMCGTLEIAPCSRVGHVFRKSTYSPFGGTSIVNK 450  
 QY 389 NLRVAEVMDEVAEYIYQRPPEYRHLSAGDVAVQKLRSLNCKSKFRMPTKIAMLPK 448  
 DB 451 NNARLAEVMDGMSSEFYTNINPGARKASAGDVSEERRALRELRCKSKFRMYLENI----- 504  
 QY 449 FYFPEVEPPAAM--GEIRNVGTGCACTDKGALGSPRLREGCVAGREBAAMNNQVTFET 506  
 DB 505 -YESQMPLYIYFGEIRNVKTNCLDTMGKSKSEKIGSSYC-HGLG-----GNQVAYIT 557  
 QY 507 WRDIRGDPQHTKKCFDAISHTSPVTLVYDCHSMKNQULMKY-RKDKTLVHPVSGSCMD 565  
 DB 558 KRHQIMSDDN-----CLDASNALGPVNLVRCHEMGNGNQEWIYDDEKTIKHAVSGNCLT 611  
 QY 566 -CESSDHRI-FMNTCNSSSLTQOWLFE 590  
 DB 612 RASBDDSTPLLRPCNYS-QQQWLMO 637

## RESULT 16

ID Q70048 PRELIMINARY; PRT; 518 AA.  
 AC Q70048;  
 DT 01-MAR-2004 (Tremblrel. 26, Created)  
 DT 01-MAR-2004 (Tremblrel. 26, Last sequence update)  
 DT 01-MAR-2004 (Tremblrel. 26, Last annotation update)  
 DE AgCP9928 (Fragment).

GN Name=agCG55772; ORFNames=ENSNANGG00000013981;  
 OS Anopheles gambiae str. PEST.  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae; Anopheles.  
 OC NCBI\_TaxID=180454;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=PEST;  
 RA Anopheles Genome Sequencing Consortium;  
 RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.  
 CC -!- CAUTION: The sequence shown here is derived from an  
 EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is  
 preliminary data.  
 DR EMBL; AAAB01008986; EAA00233.1; -;  
 DR InterPro: IPR001173; Glyco\_transf\_2;  
 DR InterPro: IPR000772; Ricin\_B\_lectin.  
 DR Pfam; PF00535; Glycos\_transf\_2; 1.  
 DR Pfam; PF00652; Ricin\_B\_lectin; 3.  
 DR PROSITE; PSS0231; RICIN\_B\_LECTIN; 1.  
 DR NON\_TER 1 1  
 FT SEQUENCE 518 AA; 58964 MW; 00F8821266B395FE CRC64;

Query Match 34.9%; Score 1144; DB 2; Length 518;  
 Best Local Similarity 45.7%; Pred. No. 5,4e-84;  
 Matches 243; Conservative 71; Mismatches 170; Indels 48; Gaps 15;

QY 88 GNGEGRPEYMTD-----AEVDAQVRENGFNIVYSDKISLNRSLPDIRHPNCNS-KRY 140  
 DB 13 GPSELGKFPVLPELPEVEKLVDDGAKNAFQYVADMIISIRTLTDPDPAWCKEEDRY 72  
 QY 141 LETLPTNTSIIIPPHNEGSSSLRTVHSVYLNPSPELVAEIVLVDPSDRHKKPLEDY 200  
 DB 73 MDLPPTSVLIICFHNEMASVLTIVHSVLDSPQVLEKIVLVDPSDMPTQQLDYF 132  
 QY 201 ALPEPVRLTKKREGIIRRMIGASATGDTITFLDSHCEAVNMLPILLDIRARRK 260  
 DB 133 LAIPKQIYVAAAREGLIRARLLGARRHATAPVLTIDSHCECTYGMLEPLIDRIARSTT 192  
 QY 261 IYCPMTDIVDHDFFRYE-TQAGDAMRGAFDEMYTKRIPRP-ELQKA--DPSDPFESP 317  
 DB 193 VCPVVIDVIDDNTMEYHYRDSGVNNGVGFDMNQLFMHAYBEEKRKHGKSPAEVWSP 252  
 QY 318 AGGLFANDRKMFHELGYPDGLIEWGEQYEISPFKVMCGRMEDIPCSRVGHIYRKVP 377  
 DB 253 AGGLFALDRVFFRLGLYDGFIDWGENLEISPFKVMCGSGLIYCSHVGHIFRRSP 312  
 QY 378 YKVPAGVS-LARLKRVAEVMDEVAEYIYQRPPEYRHLSAGDVAVQKLRSLNCK 432  
 DB 313 YKRTGVNVYIKRNSVRLAEVWMDYIYQIRGNDKGY-----GDVSRKKLREELGC 367  
 QY 433 KSFKEWPTKIAMDLPKFYFPEVEPP--AAWGEIRNVGTG--LCADTKGA--LGSPLRLB 486  
 DB 368 KSFRTYLDNI-----YPLFVPGDAVASGEVRNNGYGRRTCLDADGGRNLAKPVGLY 420  
 QY 487 GCYRGGEAAMNNQVTFETFWREDIRGDPQHTKKCFDAISHTSPVTLVYDCHSMKN 546  
 DB 421 PCNNOGNOYM--MLSKTGIRRD-----ECLDVAQ--DDVVLTPCHGSRGNOY 466  
 QY 547 WKTR-KDKTLVHPVSGSCMDSESDHRI-FMNTCNPSLSTQOWLFEHTNSTVL 597  
 DB 467 WNTSDGSHLLRHGSSDRCLAINAKNKLIMQDCAVAEQRWSPQNTDAKL 518

## RESULT 17

ID GLTI\_HUMAN STANDARD; PRT; 559 AA.  
 AC Q10472; Q9D086;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 01-OCT-2004 (Rel. 45, Last annotation update)  
 DE Polypeptide N-acetylglucosaminyltransferase 1 (EC 2.4.1.41)  
 DE (protein-UDP acetylglucosaminyltransferase 1) (UDP-  
 DE GalNAc:polypeptide N-acetylglucosaminyltransferase 1) (Polypeptide

DE GalNAc transferase 1) (GalNAc-T1) (pp-GanTrase 1) [Contains:  
 DE Polypeptide N-acetylgalactosaminyltransferase 1, soluble form].  
 GN Name=GALNT1;  
 OS Homo sapiens (Human).  
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 CC NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A. (ISOFORM 1).  
 RC TISSUE=Salivary gland;  
 RX MEDLINE=96115928; PubMed=8690719;  
 RA Meurer J.A., Naylor J.M., Baker C.A., Thomsen D.R., Homa F.L.,  
 RA Elhammer A.P.;  
 RT "CDNA cloning, expression, and chromosomal localization of a human  
 RT UDP-GalNAc:polypeptide, N-acetylgalactosaminyltransferase.";  
 RL J. Biochem. 118:568-574 (1995).  
 RN [2]  
 RP SEQUENCE FROM N.A. (ISOFORM 1), AND TISSUE SPECIFICITY.  
 RC MEDLINE=96025800; PubMed=7592619;  
 RX White T., Bennett E.P., Takio K., Soerensen T., Bonding N.,  
 RA Clausen H.;  
 RT "Purification and cDNA cloning of a human UDP-N-acetyl-alpha-D-  
 RT galactosamine:polypeptide N-acetylgalactosaminyltransferase.";  
 RL J. Biol. Chem. 270:24156-24165 (1995).  
 RN [3]  
 RP SEQUENCE FROM N.A. (ISOFORM 2).  
 RC TISSUE=PNS;  
 RX MEDLINE=22398257; PubMed=12477932; DOI=10.1073/pnas.242603899;  
 RA Straubeberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins P.S., Wagner L., Shemmen C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
 RA Datchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stedington M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Usdin T.B., Toehlyuk S., Carrinci P., Prange C.,  
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,  
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
 RA Buttefield Y.S.N., Krzywinski M.I., Skalska U., Smalhus D.E.,  
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length human  
 RT and mouse cDNA sequences.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).  
 RN [4]  
 RP SEQUENCE OF 512-559 FROM N.A. (ISOFORM 1).  
 RX MEDLINE=96285863; PubMed=8727794;  
 RA Meurer J.A., Drong R.F., Homa F.L., Slightom J.L., Elhammer A.P.;  
 RT "Organization of a human UDP-GalNAc:polypeptide, N-  
 RT acetyl-galactosaminyltransferase gene and a related processed  
 RT pseudogene.";  
 RL Glycobiology 6:231-241 (1996).  
 RN [5]  
 RP FUNCTION.  
 RX PubMed=9295285;  
 RA Wandall H.H., Haas H., Mirgorodskaya E., Kristensen A.K.,  
 RA Roepstorff P., Bennett E.P., Nielsen P.A., Hollingsworth M.A.,  
 RA Burchell J., Taylor-Papadimitriou J., Clausen H.;  
 RT "Substrate specificities of three members of the human UDP-N-acetyl-  
 RT alpha-D-galactosamine:polypeptide N-acetylgalactosaminyltransferase  
 RT family, GalNAc-T1, -T2 and -T3.";  
 RL J. Biol. Chem. 272:23503-23514 (1997).  
 RN [6]  
 RP SUBCELLULAR LOCATION.  
 RX PubMed=9394011;  
 RA Roepstorff S., White J., Wandall H.H., Olivo J.-C., Stark A.,  
 RA Bennett E.P., Whitehouse C., Berger E.G., Clausen H., Nilsson T.;  
 RT "Localization of three human polypeptide GalNAc-transferases in HeLa  
 RT cells suggests initiation of O-linked glycosylation throughout the

RT Golgi apparatus.";  
 RL J. Cell Sci. 111:45-60 (1998).  
 CC -1- FUNCTION: Catalyzes the initial reaction in O-linked  
 CC oligosaccharide biosynthesis, the transfer of an N-acetyl-D-  
 CC galactosamine residue to a serine or threonine residue on the  
 CC protein receptor. Has a broad spectrum of substrates for peptides  
 CC such as EA2, Muc5AC, Muc6, Muc7, Muc8, Muc9, Muc10, Muc11,  
 CC CATALYTIC ACTIVITY: UDP-N-acetyl-D-galactosamine + polypeptide =  
 CC UDP + N-acetyl-D-galactosaminyl-polypeptide.  
 CC -1- COFACTOR: Manganese and calcium.  
 CC -1- PATHWAY: Glycosylation.  
 CC -1- SUBCELLULAR LOCATION: Type II membrane protein. Golgi; resides  
 CC evenly across the Golgi stack. A secreted form also exists.  
 CC -1- ALTERNATIVE PRODUCTS:  
 CC Event=Alternative splicing; Named Isoforms=2;  
 CC Name=1;  
 CC IsoId=Q10472-1; Sequence=Displayed;  
 CC Name=2;  
 CC IsoId=Q10472-2; Sequence=VSP\_011200;  
 CC Note=No experimental confirmation available;  
 CC -1- TISSUE SPECIFICITY: Widely expressed. Expressed in all tissues  
 CC tested.  
 CC -1- DOMAIN: There are two conserved domains in the glycosyltransferase  
 CC region: the N-terminal domain (domain A, also called GT1 motif),  
 CC which is probably involved in manganese coordination and substrate  
 CC binding and the C-terminal domain (domain B, also called  
 CC Gal/GalNAc-T motif), which is probably involved in catalytic  
 CC reaction and UDP-gal binding (By similarity).  
 CC -1- DOMAIN: The ricin B-type lectin domain directs the glycopeptide  
 CC specificity. It is required in the glycopeptide specificity of  
 CC enzyme activity but not for activity with naked peptide  
 CC substrates, suggesting that it triggers the catalytic domain to  
 CC act on GalNAc-glycopeptide substrates (By similarity).  
 CC -1- SIMILARITY: Belongs to the glycosyltransferase family 2. GalNAc-T  
 CC subfamily.  
 CC -1- SIMILARITY: Contains 1 ricin B-type lectin domain.  
 CC -1- SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: U4514; AAC50327.1; -;  
 CC EMBL: X85019; CA59380.1; -;  
 CC EMBL: BC047746; AA47746.1; -;  
 CC EMBL: S82597; AAD14406.1; -;  
 CC PIR: JC4223; JC4223.  
 CC Genew: HGNC:4123; GALNT1.  
 CC MIM: 602273; -;  
 CC GO: 0004653; F: polypeptide N-acetylgalactosaminyltransferase. . .; TAS.  
 CC GO: 0006493; P: O-linked glycosylation; TAS.  
 CC InterPro: IPR001173; Glyco\_tran\_2.  
 CC InPro: IPR008997; Glyco\_1like.  
 CC InterPro: IPR000772; Ricin B lectin.  
 CC Pfam: PF00535; Glycos\_transf\_2; 1.  
 CC Pfam: PF00652; Ricin B lectin; 3.  
 CC SMART: SM00458; RICIN\_1.  
 CC PROSITE: PS50231; RICIN\_B\_LLECTIN; 1.  
 CC Alternative splicing: Calcium; Glycoprotein; Glycosyltransferase;  
 CC Golgi stack; Lectin; Manganese; Signal-anchor; Transferase;  
 CC Transmembrane.  
 CC PROPP 1 40  
 CC Removed in soluble polypeptide N-  
 CC acetyl-galactosaminyltransferase (By  
 CC similarity).  
 CC Polypeptide N-  
 CC acetyl-galactosaminyltransferase 1,  
 CC soluble form.  
 CC Cytoplasmic (potential).  
 CC Signal-anchor for type II membrane  
 CC protein (potential).  
 FT CHAIN 41 559  
 FT DOMAIN 1 8  
 FT TRANSMEM 9 28

FT DOMAIN 29 559 Lumenal, catalytic (Potential).

FT DOMAIN 115 225 Catalytic subdomain A.

FT DOMAIN 285 347 Catalytic subdomain B.

FT DOMAIN 429 551 Ricin B-type lectin.

FT SITE 141 141 Not glycosylated (By similarity).

FT DISULFID 442 459 By similarity.

FT DISULFID 482 497 By similarity.

FT DISULFID 523 540 By similarity.

FT CARBOHYD 95 95 N-linked (GlcNAc. .) (By similarity).

FT CARBOHYD 117 117 O-linked (Potential).

FT CARBOHYD 118 118 O-linked (Potential).

FT CARBOHYD 119 119 O-linked (Potential).

FT CARBOHYD 288 288 O-linked (Potential).

FT CARBOHYD 541 541 N-linked (GlcNAc. .) (Potential).

FT CARBOHYD 552 552 N-linked (GlcNAc. .) (By similarity).

FT VARSPLIC 106 559 Missing (in isoform 2).

FT SEQUENCE 559 AA; 64219 MW; CD68118C201E5B5 CRC64;

Query Match 34.3%; Score 1125; DB 1; Length 559;

Best Local Similarity 44.0%; Pred. No. 2,1e-82;

Matches 232; Conservative 87; Mismatches 170; Indels 38; Gaps 16;

QY GNGEGRP--YPMIDARVQAVRENGPNIVSDKISLNSLPDRHPNCSKRYLETLP 145

DB 57 GGGEMGRPVVVPKEDQKMKEMFKINDPMLASMLNRLPVRLEGCKTKYPPNLP 116

QY 146 NTSIIIFHNESGSLRTVSHVSNRSPPELVAIVLDDPSDEHKKPELEDMY-ALFP 204

DB 117 TTSVIVFHNEMAMTTLRTVSHVSNRSPRHIEIVLVDSDSDPLKRPESVYKLV 176

QY 205 SVRIITKKKBEGLRTVMGLASVATGVITFLDSHCANVMPLDRIARNKRTVCP 264

DB 177 PVHVRMRSGSLIRALKGAASVSKGVITFLDHNCECTVGMPLRLARIGHDRRTVCP 236

QY 265 MINDVDDDDRRYEQAG-DAMRGAFDWEMYYKRIPIPP--ELQKAPSDPESPWAGG 320

DB 237 IIDVSDTPEY--MAGSDMTYGGFMKLNFRMYPVQREMDRKRGGRTLPVRTPWAG 294

QY 321 LFAVDRKMFELAGYDGLIETWGEQYEISFKVMCGRMEDICSRVGHYRKVPYK 380

DB 295 LFSIDRRYFOEIGTYDAGMDWGENEISFRIMQCGTLEIVCSHVGHFRATYTF 354

QY 381 PAGVS--LARNLKRVAEVMDEVAEYIYQRRPEYRHL-SAGDAVAQKLRSSLNCKSKFWF 438

DB 355 PCGGTQIINKNNRRLAEVMDDEFKNPFYIISPVTKYDGISSRVGLRHLQCKPFSWY 414

QY 439 MTKIAMP--LPKTPPEPPRAAAGETIRNVGTGCAOTKIGALGSPRLBECVARGEEA 496

DB 415 LENTYPSQIPIRHY-----FSLGEIRNVETNOCLINMARKENKXGIFNC-HGMG--- 463

QY 497 WNNQVFTFWREDIRPGDPQHTKKFCDAISHTSPVTLVYCHSMKGNQWKYRKDK-TL 555

DB 464 --GNQVSYRANKKIRIDD-----LCLDVSKLNGPVTMLKHLKGNQMLEYIPVLTIL 515

QY 556 YHPVSGCMD-CSESDEHI-FMNTCNPSSLTQOMLFEHTNSVYLEKFP 600

DB 516 QHVNNSQCLDKATREDSQVPSIRDCN-GSRSQWLRLR--NVTLPEIF 559

RESULT 18

GLTI\_PIG STANDARD; PRT; 559 AA.

AC Q29121;

DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 01-OCT-2004 (Rel. 45, Last annotation update)

DE Polypeptide N-acetylglucosaminyltransferase 1 (EC 2.4.1.41)

DE (Protein-UDP N-acetylglucosaminyltransferase 1) (UDP-

DE GalNAc:polypeptide N-acetylglucosaminyltransferase 1) (Polypeptide

DE GalNAc:transferase 1) (GalNAc-T1) (pp-GANTase 1) [Contains:

DE Polypeptide N-acetylglucosaminyltransferase 1, soluble form].

GN Name=GALNT1;

OS Sus scrofa (Pig).

CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

CC Mammalia; Eutheria; Cetartiodactyla; Suidae; Suidae; Sus.

OX NCBI\_TaxID=9823;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Lung;

RX MEDLINE=96318021; PubMed=8748160;

RA Yoshida A., Hara T., Ikenaga H., Takeuchi M.;

RT "Cloning and expression of a porcine UDP-GalNAc: polypeptide N-

RT acetylglucosaminyl transferase."

RL Glycoconj. J. 12:824-828(1995).

CC -1- FUNCTION: Catalyzes the initial reaction in O-linked

CC oligosaccharide biosynthesis, the transfer of an N-acetyl-D-

CC galactosamine residue to a serine or threonine residue on the

CC protein receptor. Has a broad spectrum of substrates for peptides

CC such as EA2, Muc5AC, Muc6, Muc7 (By similarity).

CC -1- CATALYTIC ACTIVITY: UDP-N-acetyl-D-galactosamine + polypeptide =

CC UDP + N-acetyl-D-galactosaminyl-polypeptide.

CC -1- COFACTOR: Manganese and calcium (By similarity).

CC -1- PATHWAY: Glycosylation.

CC -1- SUBCELLULAR LOCATION: Type II membrane protein. Golgi; resides

CC evenly across the Golgi stack. A secreted form also exists (By

CC similarity).

CC -1- DOMAIN: There are two conserved domains in the glycosyltransferase

CC region: the N-terminal domain (domain A, also called GT1 motif),

CC which is probably involved in manganese coordination and substrate

CC binding and the C-terminal domain (domain B, also called

CC Gal/GalNAc-T motif), which is probably involved in catalytic

CC reaction and UDP-Gal binding (By similarity).

CC -1- DOMAIN: The ricin B-type lectin domain directs the glycopeptide

CC specificity. It is required in the glycopeptide specificity of

CC enzyme activity but not for activity with naked peptide

CC substrates, suggesting that it triggers the catalytic domain to

CC act on GalNAc-glycopeptide substrates (By similarity).

CC -1- SIMILARITY: Belongs to the glycosyltransferase family 2. GalNAc-T

CC subfamily.

CC -1- SIMILARITY: Contains 1 ricin B-type lectin domain.

CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC

DR EMBL, D85389; BAA12800.1; -;

DR InterPro; IPR001173; Glyco\_trans\_2.

DR InterPro; IPR008997; RicinB-like.

DR InterPro; IPR000772; Ricin B lectin.

DR Pfam; PF00535; Glycos\_transf\_2; 1.

DR Pfam; PF00652; Ricin B lectin; 3.

DR SMART; SM00458; RICIN\_1; RICIN\_B\_LECTIN; 1.

DR PROSITE; PS50231; RICIN\_B\_LECTIN; 1.

KW Calcium; Glycoprotein; Glycosyltransferase; Golgi stack; Lectin;

KW Manganese; Signal-anchor; Transmembrane.

FT PROPEP 1 40

FT

FT CHAIN 41 559

FT

FT Polypeptide N-

FT acetylglucosaminyltransferase 1,

FT soluble form (By similarity).

FT Cyttoplasmic (Potential).

FT Signal-anchor for type II membrane

FT protein (Potential).

FT

FT DOMAIN 29 559

FT

FT Lumenal, catalytic (Potential).

FT

FT DOMAIN 115 225

FT

FT Catalytic subdomain A.

FT

FT DOMAIN 285 347

FT

FT Catalytic subdomain B.

FT

FT DOMAIN 429 551

FT

FT Ricin B-type lectin.

FT

FT SITE 141 141

FT

FT Not glycosylated (Probable).

FT

FT DISULFID 442 459

FT

FT By similarity.

FT

FT DISULFID 482 497

FT

FT By similarity.

Query Match 34.1%, Score 1117, DB 1, Length 559;  
 Best Local Similarity 43.6%, Fred. No. 9.4e-82;  
 Matches 230; Conservative 90; Mismatches 169; Indels 38; Gaps 16;

FT	DISULFID	523	540	By similarity.
FT	CARBOHYD	95	95	N-linked (GlcNAc. . .) (By similarity).
FT	CARBOHYD	117	117	O-linked (potential).
FT	CARBOHYD	118	118	O-linked (potential).
FT	CARBOHYD	119	119	O-linked (potential).
FT	CARBOHYD	288	288	O-linked (potential).
FT	CARBOHYD	552	552	N-linked (GlcNAc. . .) (By similarity).
FT	CARBOHYD	559 AA	682CD261A51684C6 CXC64;	
SO	SEQUENCE			

Query Match 34.1%, Score 1117, DB 1, Length 559;  
 Best Local Similarity 43.6%, Fred. No. 9.4e-82;  
 Matches 230; Conservative 90; Mismatches 169; Indels 38; Gaps 16;

QY	88	NGNGRGR--YPMIDAEVDAVRENGENIYVSDKISINRSIPDIRHPCNSKRYLETLP	145
QY <td>57</td> <td>CGEGMGKGVVLPKEHQDCKMKEMFKINQNMASSEMIALNRSLPVRLEGCKTKYVPELTP</td> <td>116</td>	57	CGEGMGKGVVLPKEHQDCKMKEMFKINQNMASSEMIALNRSLPVRLEGCKTKYVPELTP	116
QY <td>146</td> <td>NTSIIIPHNMGWSLLRTVSVLNRSPPLEVAELVLDVDFSDREHLKKPLEDYW-ALFP</td> <td>204</td>	146	NTSIIIPHNMGWSLLRTVSVLNRSPPLEVAELVLDVDFSDREHLKKPLEDYW-ALFP	204
QY <td>117</td> <td>TTSVIVFHNAMWSTLRTVSHVINSRPHMLEIVLDVDSERDFLKRPLESYVKLKV</td> <td>176</td>	117	TTSVIVFHNAMWSTLRTVSHVINSRPHMLEIVLDVDSERDFLKRPLESYVKLKV	176
QY <td>205</td> <td>SVRLRTKREGLRTRMLGASVANGDYITFLDSCEANVMPLDLRIANRRTYCP</td> <td>264</td>	205	SVRLRTKREGLRTRMLGASVANGDYITFLDSCEANVMPLDLRIANRRTYCP	264
QY <td>177</td> <td>PVHVRMEQKSLRLARLGAASVQVITFLDAECCTGVLEPLRLRIKHDRKTVCP</td> <td>236</td>	177	PVHVRMEQKSLRLARLGAASVQVITFLDAECCTGVLEPLRLRIKHDRKTVCP	236
QY <td>265</td> <td>MIDVIDHDFEYETQAG-DAMRGAPEMMEYKRIPIFP---ELQKADSPDFESPVMAG</td> <td>320</td>	265	MIDVIDHDFEYETQAG-DAMRGAPEMMEYKRIPIFP---ELQKADSPDFESPVMAG	320
QY <td>237</td> <td>IIVDISDTFEY--MAGSDMTYCGGNMKNLNRWYVPREMDRKRGGDTTLVTRPTMAG</td> <td>294</td>	237	IIVDISDTFEY--MAGSDMTYCGGNMKNLNRWYVPREMDRKRGGDTTLVTRPTMAG	294
QY <td>321</td> <td>LPAVDKRMFWELGYPGLEIHWGQVYISFKVMNCGRMEDIPCSRVGHYIKRYKVPYV</td> <td>380</td>	321	LPAVDKRMFWELGYPGLEIHWGQVYISFKVMNCGRMEDIPCSRVGHYIKRYKVPYV	380
QY <td>295</td> <td>LFSIDRDYFQBIGTADAMDIMWGGENLEISFRIMCGGTLLEIVTCSHVGHFRKATPYTF</td> <td>354</td>	295	LFSIDRDYFQBIGTADAMDIMWGGENLEISFRIMCGGTLLEIVTCSHVGHFRKATPYTF	354
QY <td>381</td> <td>PAGVS--LARBKVAEYVMEDEYAYTORRPEYHLSAGVAVQKTLRSSLNCSEKFP</td> <td>438</td>	381	PAGVS--LARBKVAEYVMEDEYAYTORRPEYHLSAGVAVQKTLRSSLNCSEKFP	438
QY <td>355</td> <td>FGGIGQIINKNNRRRLAEVWMBEFTFYIISPGVTKYVGISSLGRHLKQCPFSWY</td> <td>414</td>	355	FGGIGQIINKNNRRRLAEVWMBEFTFYIISPGVTKYVGISSLGRHLKQCPFSWY	414
QY <td>439</td> <td>MTKIAMD--LRFYPPVPPAAAMGEIRNVGTLCAPIKHGALGSLRLLEGCVRGSA</td> <td>496</td>	439	MTKIAMD--LRFYPPVPPAAAMGEIRNVGTLCAPIKHGALGSLRLLEGCVRGSA	496
QY <td>415</td> <td>LENIPDSQIRPHY-----SLGEIRNVETNOCLDMARKENKVGITFNC--HMG---</td> <td>463</td>	415	LENIPDSQIRPHY-----SLGEIRNVETNOCLDMARKENKVGITFNC--HMG---	463
QY <td>497</td> <td>WNNNOVFTFWREDIRPDQHTKKCFDAISHTSPVTLVDCHSMKNQMLKRYRDK--TL</td> <td>555</td>	497	WNNNOVFTFWREDIRPDQHTKKCFDAISHTSPVTLVDCHSMKNQMLKRYRDK--TL	555
QY <td>464</td> <td>--GQVDFSYTANKERITD-----LCIDVSKLNPVYMLCKHLKQNLWEYDVKLT</td> <td>515</td>	464	--GQVDFSYTANKERITD-----LCIDVSKLNPVYMLCKHLKQNLWEYDVKLT	515
QY <td>556</td> <td>YHPVSGSMD-CSESDHRI-FMNTCPSSLTQOMFEHTNSVYLEKF</td> <td>600</td>	556	YHPVSGSMD-CSESDHRI-FMNTCPSSLTQOMFEHTNSVYLEKF	600
QY <td>516</td> <td>QVANSNOCLDKATEBDSQVPSIRDCS-GSRSQWLLR--NVTLPPIF</td> <td>559</td>	516	QVANSNOCLDKATEBDSQVPSIRDCS-GSRSQWLLR--NVTLPPIF	559

RESULT 19  
 ID GLTI\_RAT STANDARD: PRT; 559 AA.  
 AC Q10473;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 01-OCT-2004 (Rel. 45, Last annotation update)  
 DE Polypeptide N-acetylglucosaminyltransferase 1 (EC 2.4.1.41)  
 DE (Protein-UDP acetylglucosaminyltransferase 1) (UDP-  
 DE GalNAc:polypeptide N-acetylglucosaminyltransferase 1) (Polypeptide  
 DE GalNAc transferase 1) (GalNAc-T1) (dp-GalTase 1) (Contains:  
 DE Polypeptide N-acetylglucosaminyltransferase 1, soluble form).  
 GN Name=GalnT1;  
 OS Rattus norvegicus (Rat).  
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 CC NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STEIN-Sprague-Dawley; TISSUE=Sublingual gland;  
 RX MEDLINE=96318029; PubMed=8748168;  
 RA Hagen F., Gregorie C.A., Tabak L.A.;  
 RT "Cloning and sequence homology of a rat UDP-GalNAc:polypeptide N-

RT acetylglucosaminyltransferase.";  
 RL Glycoconj. J. 12:901-909(1995).  
 RN [2]  
 RP MUTAGENESIS OF CYS-106, CYS-212, CYS-214, CYS-235, CYS-330, CYS-339  
 RP AND CYS-408.  
 RX PubMed=1219709;  
 RA Tanno M., Toba S., Kozdy F.J., Elhammer A.P., Kurosaka A.;  
 RT "Identification of two cysteine residues involved in the binding of  
 RT UDP-GalNAc to UDP-GalNAc:polypeptide N-acetylglucosaminyltransferase  
 RT 1 (GalNAc-T1)."  
 RL Eur. J. Biochem. 269:4308-4316(2002).  
 RN [3]  
 RP MUTAGENESIS OF CYS-442, CYS-459, CYS-482, CYS-497, CYS-523, CYS-540;  
 RP ASP-444; GLY-455; PHE-457; PHE-468; ASN-465; GLN-466; ASP-484 AND  
 RP ASP-525.  
 RX PubMed=12364335; DOI=10.1074/jbc.M207369200;  
 RA Tanno M., Saeki A., Kozdy F.J., Elhammer A.P., Kurosaka A.;  
 RT "The lectin domain of UDP-GalNAc:polypeptide N-  
 RT acetylglucosaminyltransferase 1 is involved in O-glycosylation of a  
 RT polypeptide with multiple acceptor sites."  
 RL J. Biol. Chem. 277:47088-47096(2002).  
 RN [4]  
 RP MUTAGENESIS OF ASP-444, ASP-484 AND ASP-525.  
 RX PubMed=12419318;  
 RA Tanno M., Kozdy F.J., Elhammer A.P., Kurosaka A.;  
 RT "Function of the lectin domain of polypeptide N-  
 RT acetylglucosaminyltransferase 1."  
 RL Biochem. Biophys. Res. Commun. 298:755-759(2002).  
 CC -1- FUNCTION: Catalyzes the initial reaction in O-linked  
 CC oligosaccharide biosynthesis, the transfer of an N-acetyl-D-  
 CC galactosamine residue to a serine or threonine residue on the  
 CC protein receptor. Has a broad spectrum of substrates for peptides  
 CC such as EA2, Muc5AC, Muc6, Muc7, Muc8 and Muc9.  
 CC -1- CATALYTIC ACTIVITY: UDP-N-acetyl-D-galactosamine + polypeptide =  
 CC UDP + N-acetyl-D-galactosaminyl-polypeptide.  
 CC -1- COFACTOR: Manganese and calcium.  
 CC -1- PATHWAY: Glycosylation.  
 CC -1- SUBCELLULAR LOCATION: Type II membrane protein. Golgi; resides  
 CC evenly across the Golgi stack. A secreted form also exists (By  
 CC similarity).  
 CC -1- TISSUE SPECIFICITY: Heart, brain, spleen, liver, skeletal muscle  
 CC and kidney.  
 CC -1- DOMAIN: There are two conserved domains in the glycosyltransferase  
 CC region: the N-terminal domain (domain A, also called G1 motif),  
 CC which is probably involved in manganese coordination and substrate  
 CC binding and the C-terminal domain (domain B, also called  
 CC Gal/GalNAc-T motif), which is probably involved in catalytic  
 CC reaction and UDP-Gal binding.  
 CC -1- DOMAIN: The rctin B-type lectin domain directs the glycopeptide  
 CC specificity. It is required in the glycopeptide specificity of  
 CC enzyme activity but not for activity with naked peptide  
 CC substrates, suggesting that it triggers the catalytic domain to  
 CC act on GalNAc-glycopeptide substrates.  
 CC -1- SIMILARITY: Belongs to the glycosyltransferase family 2. GalNAc-T  
 CC subfamily.  
 CC -1- SIMILARITY: Contains 1 rctin B-type lectin domain.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.ebi.ac.uk/announcements/>  
 CC or send an email to [license@ebi.ac.uk](mailto:license@ebi.ac.uk)).  
 CC -----  
 CC EMBL, U35890; AAC52511.1; -.  
 CC RCD, 620358; GalnT1.  
 CC InterPro: IPR001173; Glyco trans 2.  
 CC InterPro: IPR008997; RctinB like.  
 CC InterPro: IPR000772; RctinB lectin.  
 CC Pfam: PF00535; Glycosyl transferase 2, 1.  
 CC Pfam: PF00652; RctinB lectin; 3.  
 CC SMART, SM00458; RCTIN, 1.

DR PROSITE; PS50231; RICIN B LECTIN; 1.  
 KM Calcium; Glycoprotein; Glycosyltransferase; Golgi stack; Lectin;  
 FT Mangano; Signal-anchor; Transmembrane.  
 FT PROPEP 1 40 Removed in soluble polypeptide N-acetylglucosaminyltransferase.  
 FT CHAIN 41 559 acetylglucosaminyltransferase 1, soluble form.  
 FT DOMAIN 1 8 Cytoplasmic (potential).  
 FT TRANSMEM 9 28 Signal-anchor for type II membrane protein (potential).  
 FT DOMAIN 29 559 Luminal, catalytic (potential).  
 FT DOMAIN 115 225 Catalytic subdomain A.  
 FT DOMAIN 285 347 Catalytic subdomain B.  
 FT DOMAIN 429 551 Ricin B-type lectin.  
 FT SITE 141 141 Not glycosylated (by similarity).  
 FT DISULFID 442 459 By similarity.  
 FT DISULFID 482 497 By similarity.  
 FT CARBOHYD 523 540 By similarity.  
 FT CARBOHYD 95 95 N-linked (GlcNAc...) (by similarity).  
 FT CARBOHYD 117 117 O-linked (potential).  
 FT CARBOHYD 118 118 O-linked (potential).  
 FT CARBOHYD 119 119 O-linked (potential).  
 FT CARBOHYD 288 288 O-linked (potential).  
 FT CARBOHYD 552 552 N-linked (GlcNAc...) (by similarity).  
 FT MUTAGEN 106 106 C->A: Loss of function.  
 FT MUTAGEN 212 212 C->A: Loss of function due to absence of interaction between UDP moiety and UDP-GalNAc.  
 FT MUTAGEN 214 214 C->A: Loss of function due to absence of interaction between UDP moiety and UDP-GalNAc.  
 FT MUTAGEN 235 235 C->A: No effect.  
 FT MUTAGEN 330 330 C->A: Loss of function.  
 FT MUTAGEN 339 339 C->A: Loss of function.  
 FT MUTAGEN 408 408 C->A: Loss of function.  
 FT MUTAGEN 442 442 C->A: Loss of function.  
 FT MUTAGEN 444 444 D->A: Induces a strong decrease in activity; Loss of function; when associated with A-484 and A-525.  
 FT MUTAGEN 455 455 F->Q: Induces a decrease in activity.  
 FT MUTAGEN 457 457 F->A: Little or no effect.  
 FT MUTAGEN 459 459 C->A: Loss of function.  
 FT MUTAGEN 465 465 N->A: Little or no effect.  
 FT MUTAGEN 466 466 Q->A: Induces a decrease in activity.  
 FT MUTAGEN 468 468 F->W/Y: Little or no effect.  
 FT MUTAGEN 482 482 C->A: Loss of function.  
 FT MUTAGEN 484 484 D->A: Loss of function; when associated with A-444 and A-525.  
 FT MUTAGEN 497 497 C->A: Loss of function.  
 FT MUTAGEN 523 523 C->A: Loss of function.  
 FT MUTAGEN 525 525 D->A: Loss of function; when associated with A-444 and A-484.  
 FT MUTAGEN 540 540 C->A: Loss of function.  
 FT SEQUENCE 559 AA; 64229 MM; SE36A95D9422C853 CRC64;  
 Query Match 34.0%; Score 116; DB 1; Length 559;  
 Best Local Similarity 43.6%; Pred. No. 1,1e-81;  
 Matches 230; Conservative 86; Mismatches 171; Indels 38; Gaps 16;

QY 265 MIDVIDHDDPVEYTOAG-DAMRGAFDWEYTKRIPPE---ELQKADPSDFESPVMAG 320  
 Db 237 IIDVIDSDTREV--MAGSDMTYGGFNKKNRPVQREMRRKGRITPVRPTTAAAG 294  
 QY 321 LPAVDRKMFELGYDGLIWIWGEBOYEISPKYMMCGRMEDIPCSRVGHYKRYVYKV 380  
 Db 295 LFSIDRDYFQIIGYDAGMDIWGENLEISFRIWQCGTLEIYVCSHVGHVFRATPYTF 354  
 QY 381 PAGVS--LAARKRVAVVMDVEAYEYTORPREYRHLSAGDAVQKLRSLNCKSRWF 438  
 Db 355 PGGGQIINKNNRLAEVMDDEFKNFFYIISPGVTKYDYDISRVLRLHLDQCKPSWY 414  
 QY 439 MTKIAMD--LPKFPPEPPAAMGEIRNVTGCLADTKGALGSPRLREGCVGRBAA 496  
 Db 415 LENIYPSQIPRH-----FSLGEIRNVEITNOCLDNMARKENKYGITRNC-HGMG--- 463  
 QY 497 WNNQVETFTWRDIRPDPOHTKKPCFPAISHTSPVTLDYCHSMKNQOLMKYRKDK-TL 555  
 Db 464 --GNQVSYTANKEIRND-----LCDVSKLNGPVTMLKCHHLKGNQLMVEYDPAVLT 515  
 QY 556 YHPVSGCMD-CSESDHRT-FNATCNPSSLTOQLFEHTNSTYLEKF 600  
 Db 516 QHVNNSQCLDKATEBDSQVPSIRDCT-GSRSQOMLR--NVTLPKELF 559  
 RESULT 20  
 GLTI\_BOVIN STANDARD; PRT; 559 AA.  
 ID GLTI\_BOVIN  
 AC 007537;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 01-OCT-2004 (Rel. 45, Last annotation update)  
 DE Polypeptide N-acetylglucosaminyltransferase 1 (EC 2.4.1.11)  
 DE (Protein-UDP acetylglucosaminyltransferase 1) (UDP-  
 DE GalNAc:polypeptide N-acetylglucosaminyltransferase 1) (Polypeptide  
 DE GalNAc transferase 1) [GalNAc-T1] (pp-GalNAse 1) [Contains:  
 DE Polypeptide N-acetylglucosaminyltransferase 1, soluble form].  
 GN Name=GLNT1;  
 OS Bos taurus (Bovine).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
 OC Bovinae; Bos.  
 OX NCBI\_TaxID=9913;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 41-74.  
 RC TISSUE=Colostrum, and Intestine;  
 RX MEDLINE=93286099; PubMed=7685345;  
 RA Homa F.L., Hollander T., Lehman D.J., Thomsen D.R., Elhammer A.P.;  
 RT "Isolation and expression of a cDNA clone encoding a bovine UDP-  
 RT GalNAc:polypeptide N-acetylglucosaminyltransferase.";  
 RL J. Biol. Chem. 268:12609-12616 (1993).  
 RN [2]  
 RP SEQUENCE OF 41-559 FROM N.A., AND PARTIAL SEQUENCE.  
 RC TISSUE=Colostrum, and Placenta;  
 RX MEDLINE=93366815; PubMed=8360184;  
 RA Hagen F.K., van Wuyckhuysen B., Tabak L.A.;  
 RT "Purification, cloning, and expression of a bovine UDP-GalNAc:  
 RT polypeptide N-acetylglucosaminyltransferase.";  
 RL J. Biol. Chem. 268:18960-18965 (1993).  
 RN [3]  
 RP MUTAGENESIS OF ASN-95, HIS-125, HIS-137, ASN-141, HIS-146, HIS-179;  
 RP HIS-211, HIS-228, HIS-341, HIS-344, HIS-404, HIS-427, HIS-460;  
 RP HIS-498, HIS-499, HIS-517 AND ASN-552.  
 RX PubMed=9359852;  
 RA Wragg S., Hagen F.K., Tabak L.A.;  
 RT "Identification of essential histidine residues in UDP-N-acetyl-D-  
 RT galactosamine:polypeptide N-acetylglucosaminyltransferase-T1.";  
 RL Biochem. J. 328:193-197 (1997).  
 CC -FUNCTION: Catalyzes the initial reaction in O-linked  
 CC oligosaccharide biosynthesis, the transfer of an N-acetyl-D-  
 CC galactosamine residue to a serine or threonine residue on the  
 CC protein receptor. Has a broad spectrum of substrates for peptides  
 CC such as E2, Muc5AC, Muc1a, Muc1b and Muc7.

CC -1- CATALYTIC ACTIVITY: UDP-N-acetyl-D-galactosamine + polypeptide =  
 CC UDP + N-acetyl-D-galactosaminyl-polypeptide.  
 CC -1- COFACTOR: Manganese and calcium.  
 CC -1- PATHWAY: Glycosylation.  
 CC -1- SUBCELLULAR LOCATION: Type II membrane protein. Golgi; resides  
 CC evenly across the Golgi stack. A secreted form also exists (By  
 CC similarity).  
 CC -1- TISSUE SPECIFICITY: Colostrum contains a soluble form.  
 CC -1- DOMAIN: There are two conserved domains in the glycosyltransferase  
 CC region: the N-terminal domain (domain A, also called GT1 motif),  
 CC which is probably involved in manganese coordination and substrate  
 CC binding and the C-terminal domain (domain B, also called  
 CC Gal/galNac-T motif), which is probably involved in catalytic  
 CC reaction and UDP-Gal binding (By similarity).  
 CC -1- DOMAIN: The ricin B-type lectin domain directs the glycopeptide  
 CC specificity. It is required in the glycopeptide specificity of  
 CC enzyme activity but not for activity with naked peptide  
 CC substrates, suggesting that it triggers the catalytic domain to  
 CC act on GalNac-glycopeptide substrates (By similarity).  
 CC -1- SIMILARITY: Belongs to the glycosyltransferase family 2. GalNac-T  
 CC subfamily.  
 CC -1- SIMILARITY: Contains 1 ricin B-type lectin domain.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.ebi.ac.uk/announcements>  
 CC or send an email to [license@ebi.ac.uk](mailto:license@ebi.ac.uk)).  
 CC -----  
 DR EMBL: L07780; AAA30532.1; -;  
 DR EMBL: L17437; AAA68489.1; -;  
 DR PIR: A45987; A45987.  
 DR InterPro: IPR001173; Glyco\_trans\_2.  
 DR InterPro: IPR008997; RicinB\_like.  
 DR InterPro: IPR000772; Ricin\_B\_lectin.  
 DR Pfam: PF00535; Glycos\_transf\_2; 1.  
 DR Pfam: PF00652; Ricin\_B\_lectin; 3.  
 DR SMART: SM00458; RICIN\_B\_1.  
 DR PROSITE: PS50231; RICIN\_B\_LECTIN; 1.  
 KM Calcium; Direct protein sequencing; Glycoprotein; Glycosyltransferase;  
 KM Golgi stack; Lectin; Manganese; Signal-anchor; Transferase;  
 KM Transmembrane.  
 FT PROPEP 1 40 Removed in soluble polypeptide N-  
 FT CHAIN 41 559 acetylgalactosaminyltransferase.  
 FT DOMAIN 1 8 Polypeptide N-  
 FT TRANSMEM 9 28 acetylgalactosaminyltransferase 1,  
 FT soluble form.  
 FT DOMAIN 1 8 Cytoplasmic (Potential).  
 FT TRANSMEM 9 28 Signal-anchor for type II membrane  
 FT protein (Potential).  
 FT DOMAIN 1 8 Catalytic subdomain A.  
 FT TRANSMEM 9 28 Catalytic subdomain B.  
 FT SITE 141 141 Not glycosylated (Probable).  
 FT DISULFID 442 459 By similarity.  
 FT DISULFID 482 497 By similarity.  
 FT DISULFID 523 540 By similarity.  
 FT CARBOHYD 141 147 O-linked (GlcNAc... ) (Probable).  
 FT CARBOHYD 117 117 O-linked (Potential).  
 FT CARBOHYD 118 118 O-linked (Potential).  
 FT CARBOHYD 119 119 O-linked (Potential).  
 FT CARBOHYD 288 288 O-linked (Potential).  
 FT CARBOHYD 552 552 N-linked (GlcNAc... ) (Probable).  
 FT MUTAGEN 95 95 N->Q: Induces decrease in glycosylation.  
 FT MUTAGEN 125 125 H->A: Induces a strong decrease in  
 FT activity.  
 FT MUTAGEN 137 137 H->A: No effect.  
 FT MUTAGEN 141 141 N->Q: No effect.  
 FT MUTAGEN 146 146 H->A: No effect.  
 FT MUTAGEN 179 179 H->A: No effect.  
 FT MUTAGEN 211 211 H->A: Loss of function.

FT MUTAGEN 228 228 H->A: No effect.  
 FT MUTAGEN 341 341 H->A: Induces a strong decrease in  
 FT activity.  
 FT MUTAGEN 344 344 H->A: Loss of function.  
 FT MUTAGEN 404 404 H->A: No effect.  
 FT MUTAGEN 427 427 H->A: No effect.  
 FT MUTAGEN 460 460 H->A: No effect.  
 FT MUTAGEN 498 498 H->A: No effect.  
 FT MUTAGEN 499 499 H->A: No effect.  
 FT MUTAGEN 517 517 H->A: No effect.  
 FT MUTAGEN 552 552 N->Q: Induces decrease in glycosylation.  
 FT SEQUENCE 559 AA; 64192 MW; B3E538C4DB569B40 CRC64;  
 SQ  
 Query Match 34.0%; Score 1115; DB 1; Length 559;  
 Best Local Similarity 43.6%; Pred. No. 1,4e-81;  
 Matches 230; Conservative 89; Mismatches 170; Indels 38; Gaps 16;  
 QY 88 GNGEQGRP--YPMTDARVDQATRENGFNIVSDKISLNSLPPIRPNCKSKRYLETLP 145  
 DB 57 GPGEMGRPVVLPKEDQKKMKEMPKINQFNLMASMIANSLPVRLEGCTKYVPNNLP 116  
 QY 146 NTSIIIFHNHGSGLRTVHSVLRSPPELVAVIYVDDPSDRHLKKPLEDYM-ALFP 204  
 DB 117 TTSVIVIFHNHAGSLRTVHSVLRSPPELVAVIYVDDPSDRHLKKPLEDYVKKLV 176  
 QY 205 SVRIIRTKKRGELIRITMLGASVATGDIITFLDSCHANYMWLPPLDIRIARNKTIYCP 264  
 DB 177 PVHIVRMQRGGLIRARLKGAAVSGQVITFLDHCCTGWMLEPLARIKHDRKTVVCP 236  
 QY 265 MIDVIDHDDPFIETQAG-DAMRGAFDMEYTKRIPIPP--ELQKAPSPDFEPVWAGG 320  
 DB 237 IIDVISDITPEY--MAQSDMTYGGFNWKLNRWYVPVQREMDRKGRTLPVRIPTWAGG 294  
 QY 321 LPAVDKRMFELGSDYDGLBTWGEQYEISPKVMCGRMEDIPCSRGHVIRKYVYKV 380  
 DB 295 LFSIDRDYFQETIGYDAGMDWGSSENEISRIQCSTLEIYVCHVGHFRATPTTF 354  
 QY 381 PAGVS--LARNLKRAVEMDEVAYIYORRPEYRHLISAGDAVQKLRSSLNCKSPKWF 438  
 DB 355 PGGTGQILINKNRRLAEVWMDPEFNFIISPGVTKYVDGISRLGLRHLQCRPSWY 414  
 QY 439 MTKLAMP-LPKFPPVPPPAANGELIRNVGTGICADTKIGALGSLRLBECVARGEEA 496  
 DB 415 LENIYPPSQIPRHY-----FSLGRIINWETNOCILNMARKENKXGIFNC-HGMG--- 463  
 QY 497 WNNQVFTFTWRREDIRPGDPHTKKCFEPAISHTSPVLYDCHSMKGNOLMKYRKDK-TL 555  
 DB 464 --GNQVSYTNKKEIRTD-----LCLDVSKLNGPVTMLCKHLKGNQMLEYPAVLTL 515  
 QY 556 YHPVSGSCMD-CSESDHRI-FMNTCNPSLSLTQWLFEHTNSTVLEKFP 600  
 DB 516 QHVNNSNQCLDKATDDBDSQVPSIRDCS-GSRQQLLR--NVTLPRIIF 559  
 RESULT 21  
 GLTI\_MOUSE STANDARD; PRT; 559 AA.  
 ID GLTI\_MOUSE  
 AC 008912; OTTNDL;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 01-OCT-2004 (Rel. 45, Last annotation update)  
 DE Polypeptide N-acetylglucosaminyltransferase 1 (EC 2.4.1.41)  
 DE (Protein-UDP N-acetylglucosaminyltransferase 1) (UDP-  
 DE GalNac:polypeptide N-acetylglucosaminyltransferase 1) (Polypeptide  
 DE GalNac transferase 1) (GalNac-T) (pp-GaNTase 1) [Contains:  
 DE Polypeptide N-acetylglucosaminyltransferase 1, soluble form].  
 GN Name=Galnt1;  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 SEQUENCE FROM N.A., FUNCTION, AND TISSUE SPECIFICITY.



```

FT MUTAGEN 460 460 H->K,N: Induces a slight decrease in
FT MUTAGEN 464 464 activity.
FT MUTAGEN 465 465 G->A: Induces a slight decrease in
FT MUTAGEN 465 465 activity.
FT MUTAGEN 465 465 NO->QN: Induces a decrease in activity.
FT MUTAGEN 465 465 N->A: Little or no effect; induces a
FT MUTAGEN 465 465 slight decrease in activity when
FT MUTAGEN 465 465 associated with A-466.
FT MUTAGEN 466 466 O->A: Little or no effect; induces a
FT MUTAGEN 466 466 slight decrease in activity when
FT MUTAGEN 466 466 associated with A-465.
FT MUTAGEN 466 466 I->T (in Ref. 2).
FT CONFLICT 81 81
FT SEQUENCE 559 AA; 64255 MW; ACSAB55D91F83E4 CRC64;
SQ
Query Match 34.0%; Score 1115; DB 1; Length 559;
Best Local Similarity 43.6%; Pred. NO. 1.4e-81;
Matches 230; Conservative 87; Mismatches 172; Indels 38; Gaps 16;
QY 88 GNGEGRD--YPMTDARVDQAYRENGFNIVSDKISLNSLPDIRHPCNSKRYLETLP 145
DB 57 GPGEMGKRVIPKEDQEKEMFKINQENLMASEMIALNRSILPVRLEGCKTKVYPDLF 116
QY 146 NTSIIIPHNCGSGLRTHTSVLNRSPELVAVIVYDPSDRHLKKEPVEDM-ALFP 204
DB 117 TTSVIVFHNMAWSTLRTVHSVNRSPHMEIVLVDASERDFLRPSYVKKLKV 176
QY 205 SVRIKRTKREGILRTMLGASVATGDTYTFDISHCEANVMWLPDLRIARNKTIYCP 264
DB 177 PVAVIRMEQSGSLRARKGAASVNGVITFLDACECTAGLELRLIRINDRTIVCP 236
QY 265 MIDVIDHDDFRYETQAG-DAMRGAFDWMYKRIPIRP--ELQKADSPDFESPVNAAG 320
DB 237 IIDVIDSDTFEY--MAGSDMTYGGHNMKINFRWYVPREMDRKRKDRTLFVRPTWAGG 294
QY 321 LFAVDKRFMELGCDPLCETLWGGQVYSFKVMCGRMEDIPSCRGHYIKTVKPVKV 380
DB 295 LFSIDRDYFOEIGYDAGMDIWGGENLEISFRIMCGGTLTIVTSHGHFRKATPYTF 354
QY 381 PAGRS--LARIKRAVWMDVAVIYQRRPEYHLGAGVAVOVKLRSLCKSKFNF 438
DB 355 PGGGQIINKNNRRLAEVWMDPKNFYIISBQYTKVDYGDSSRLGLRKLCKCFPNWY 414
QY 439 WTKIAMD-LPKFYPPVPPAAWGEIRNVGTGLCADTRGALGSPFLREGCVRGRGAA 496
DB 415 LENIYPSDQIDPRHY-----FSLGEIRNVETNQCLDMARKENKVGIFNC-HGMG--- 463
QY 497 WNNNOVFTFTWRBDIRPDPOHTKKFCDDAISHTSPVTLVYCHSKMGKQMLKRYKDK-TL 555
DB 464 --GNQVSESYTANKERTD-----LCIDVSKLNGPVTMLKCHLKGQMLWEYDPVKLT 515
QY 556 YHPVSGCMD-CSESDHRI-FWNTGNPSLTQQWLFHTNSTVLEKF 600
DB 516 QHVNNSQCLDKATBEDSQVPSIRCT-GSRQSQMLLR--NTLPEIF 559

```

```

RX MEDLINE=98192620; PubMed=9525933;
RA Hagen F.K., Nehke K.;
RT "cDNA cloning and expression of a family of UDP-N-acetyl-D-
RT galactosamine:polypeptide N-acetylgalactosaminyltransferase sequence
RT homologs from Ctenorhabditis elegans."
RL J. Biol. Chem. 273:8268-8277(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Stoneking T., Wohlmann P., Lennox S.;
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: Probable glycopeptide transferase involved in O-linked
CC oligosaccharide biosynthesis. Glycopeptide transferases catalyze
CC the transfer of an N-acetyl-D-galactosamine residue to an already
CC glycosylated peptide (By similarity). In contrast to other members
CC of the family, it does not act as a peptide transferase that
CC transfers GalNAc onto serine or threonine residue on peptides that
CC have been tested. Some peptide transferase activity is however not
CC excluded, considering that its appropriate peptide substrate may
CC remain unidentified.
CC -1- COFACTOR: Manganese and calcium (By similarity).
CC -1- PATHWAY: Glycosylation.
CC -1- SUBCELLULAR LOCATION: Type II membrane protein. Golgi (By
CC similarity).
CC -1- DOMAIN: There are two conserved domains in the glycosyltransferase
CC region: the N-terminal domain (domain A, also called G1 motif),
CC which is probably involved in manganese coordination and substrate
CC binding and the C-terminal domain (domain B, also called
CC Gal/GalNAc-T motif), which is probably involved in catalytic
CC reaction and UDP-Gal binding (By similarity).
CC -1- DOMAIN: The ricin B-type lectin domain binds to GalNAc and
CC contributes to the glycopeptide specificity (By similarity).
CC -1- SIMILARITY: Belongs to the glycosyltransferase family 2. GalNAc-T
CC subfamily.
CC
CC -1- SIMILARITY: Contains 1 ricin B-type lectin domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AF031841; AAC13677.1; -.
CC DR EMBL; AC006674; AAF60620.1; -.
CC DR PIR; T42251, T42251.
CC DR WormPep; Y46H3A.6; CE24309.
CC DR InterPro; IPR001173; Glyco_transf_2.
CC DR InterPro; IPR008997; RicinB_1like.
CC DR InterPro; IPR000772; RicinB_lectin.
CC DR Pfam; PF00535; Glycos_transf_2; 1.
CC DR Pfam; PF00652; RicinB_lectin; 3.
CC DR SMART; SM00458; RICIN_1.
CC DR PROSITE; PS50231; RICIN_B_LECTIN; 1.
CC DR Calcium; Glycosyltransferase; Golgi stack; Lectin; Manganese;
CC Signal-anchor; Transferase; Transmembrane.
CC KW DOMAIN 1 20 Cytoplasmic (Potential).
CC FT TRANSMEM 21 38 Signal-anchor for type II membrane
CC FT DOMAIN 39 601 protein (Potential).
CC FT DOMAIN 155 265 Luminal (Potential).
CC FT DOMAIN 328 390 Catalytic subdomain A.
CC FT DOMAIN 477 595 Ricin B-type lectin.
CC FT DISULFID 490 506 By similarity.
CC FT DISULFID 529 542 By similarity.
CC FT DISULFID 568 583 By similarity.
CC FT CARBOHYD 135 135 N-linked (GlcNAc....) (Potential).
CC SQ SEQUENCE 601 AA; 68108 MW; 1A7E744C93BDCD87 CRC64;

```

```

Query Match 34.0%; Score 1113; DB 1; Length 601;
Best Local Similarity 42.0%; Pred. NO. 2.2e-81;
Matches 226; Conservative 90; Mismatches 186; Indels 36; Gaps 13;

```

```

QY 68 GQKXMDHDKAIRRDQVRNGEGRRPYMTDAERV---DQAYRNGFNITYSPKISLN 124
DB 79 GSELGNEYEPKPEIPSNQ---PGEHGKVPVTDDEGMAAGAAEKEFPNTYISMTISMN 135
QY 125 RSLPDRIHPNCSKRYLETLPNTSIIIPHNMGSSLLRTVHSVLRNRPPELVAEIVLD 184
DB 136 RTIIDIRECKGMDYPRKLPVSVVVVFHNEGMPPLRTVHSVLRNRPPELVAEIVLD 195
QY 185 DFSDEHKKKPLEDMALP-DSVRLRTKREGLRTVHMGLASVATGDTVITFLDSHCAN 243
DB 196 DSDSKPHLKEKLDKVTFRNGKVIYVTRTEGEGILNARSIGAKISTGSEVFLDANCEVN 255
QY 244 VNMPLPDLIRIARRKTIIVCPMIDVIDHDDFRYEQAGDA---MRGADNEMKYKRIPI 300
DB 256 TNNMLPPLAIPKRNKRVTVTVIIGDISNSWEYSVYSPNANHSGIGEMGLKKEQIT 315
QY 301 PE--LQKADPSDPESPVMAGGLFVADRKFMEGLGYDGLIWMGEGYISFKVMWCGG 358
DB 316 ERETNRKHNQPRSPFTHAGGLPAINLMKEGLYBGLQINGEGYEISFKIMCGG 375
QY 359 RMEDIPCSRGVGHYRKVPY---KVPAGVSLARNIKRVAEVMDEYAEYIYQREPEYHL 415
DB 376 GIVFVPCSHVGHVYRSHMPSFGKFSKRPVISINMMRVVKTWMDYDSKYLTRBPQATNV 435
QY 416 SAGDVAVOKLRSSLNCSFKMFMKTIAMDLPKTYPRPPEPAAANGELIRNVTGLCADTK 475
DB 436 NPGDISAQALARDKIQCKSFPMYENVAIVDLKSPML-PRNDWGEARNATGICLD-R 493
QY 476 HGAIGSPRLREGCVRGGEAAMNNQVFTFTWREDIRGDDPOHTKFCFDHISHTSPVTL 535
DB 494 MGGLPGEAGATGC-----HGCGNQLRLNVOGOMAGB-----KCLTA---NGIRI 537
QY 536 YDCHSMKG-NQLMKY-RKDKTYHPVSGSCMDCSBDHRIEMNTCNBSLTLQOMLFE 590
DB 538 QANHCVKGTVNGFMSYDKTKQIHSQKQCIIVSSESVTLQCTEDNRQKQVWK 595

```

## RESULT 23

Q6GM51 PRELIMINARY; PRT; 556 AA.

```

ID 06GM51; AC 06GM51; DT 05-JUN-2004 (Tremblrel. 27, Created)
DT 05-JUN-2004 (Tremblrel. 27, Last sequence update)
DE Hypothetical protein.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Butelacostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidea; Pipidae;
OC Xenopodinae; Xenopus.
OC NCBI_Taxid=8355;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=EYE;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stiepleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Schetz T.E.,
RA Brownstein W.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Rana S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bock S.A., McWeney P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Buterfield Y.S.,
RA Krzywinski M.I., Skalske U., Smallie D.E., Scherch A., Schein J.R.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human

```

```

RT and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=EYE;
RX MEDLINE=22341132; PubMed=12454917;
RA Klein S.L., Strausberg R.L., Wagner L., Pontius J., Clifton S.W.,
RA Richardson P.,
RT "Genetic and genomic tools for Xenopus research: The NIH Xenopus
RT initiative.";
RL Dev. Dyn. 225:384-391(2002).
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=EYE;
RA Klein S., Strausberg R.,
RL Submitted (JUN-2004) to the EMBL/Genbank/DBJ databases.
DR EMBL; BC074234; AAH74234.1;
DR InterPro; IPR001173; Glyco trans. 2.
DR InterPro; IPR008997; RicinB like.
DR InterPro; IPR00772; Ricin_B_lectin.
DR Pfam; PF00535; Glycof. transF. 2;
DR Pfam; PF00652; Ricin_B_lectin; 3.
DR SMART; SM00458; RICIN; 1.
DR PROSITE; PS0231; RICIN_B_LECTIN; 1.
KW Hypothetical protein.
SQ SEQUENCE 556 AA; 63908 MW; 3A9FF34A39AB655B CRC64;

```

```

Query Match 33.9%; Score 1112.5; DB 2; Length 556;
Best Local Similarity 42.8%; Pred. No. 2.2e-81;
Matches 234; Conservative 91; Mismatches 181; Indels 41; Gaps 16;

```

```

QY 64 FLGSGQKXMDHDKAIRRDQVRNGEGRRPYMTDAERV---DQAYRNGFNITYSPKISLN 116
DB 27 YFSECNCDKDRKBSLLPALRAVIRSBEGBEGKAVIIPKDOEKKEKLFKINQFLM 86
QY 117 VSDKISLNRSLPDIRHNCNKRKYLETLPNTSIIIPHNMGSSLLRTVHSVLRNRPPELV 176
DB 87 ASDLIANRSLPDIRHNCNKRKYLETLPNTSIIIPHNMGSSLLRTVHSVLRNRPPELV 146
QY 177 VAEIVLVDDPSDEHKKKPLEDMALP-DSVRLRTKREGLRTVHMGLASVATGDTVITF 235
DB 147 ISEIIVDDASERFLTPTPLENYKHEVAVKILRMSORGLIARISGANVAAGKITT 206
QY 236 LDHSGEANVMPLPDLIRIARRKTIIVCPMIDVIDHDDFRYEQAG-DANRGAPEWEMY 294
DB 207 LDAHCECTFGWLBELRLARIKEDRKTVCPDIIDVISDPTFEY-MAGSDMTYGGFNMKINF 264
QY 295 KRIRIPP---ELQKADSDPESPVMAGGLFVADRKFMEGLGYDGLIWMGEGYISFK 351
DB 265 RMYVPVPREMDRRKGDRTLPVTRPTMAGGLPSIDKTYFEEGLTYSQMDIWMGLENLWSF 324
QY 352 KVMWCGRMEDIPCSRGVGHYRKVPYPRVAGVS--LARNIKRVAEVMDEYAEYIYQRR 409
DB 325 RIMWCGSLLEIVCSHGVRKATPYTFPGGTHVINKNNRLAEVWMDPFOFFYIIS 384
QY 410 PEYHLSAGDVAVOKLRSSLNCSFKMFMKTIAMDLPKTYPRPPEPAA-AMGBIRNVG 467
DB 385 PGVVAVVYGVDSSEKAKARENKCPFMWYLETV-----YPSQIPRRYPSLSEIRNVE 437
QY 468 TGLCADTKHAGLSPLLEGCVRGGEAAMNNQVFTFTWREDIRGDDPOHTKFCFDPAI 527
DB 438 TNOCLMDNKRKREKNEKVGIFNC-HGMG-----GNQVFSYTDKELRTDD-----LCIDVS 485
QY 528 SHTSPVTLTYDCHSMKGNQMLKYRKDK-TLYHPVSGSCMDCSBDHRI-F-MNTCNPSL 584
DB 486 RINPFPVIMLCKHNRGNQWBYDAEHLILRHINSNGLDDEADDDKXVPTIKECN-GSRS 544
QY 585 QQWLFEH 591
DB 545 QQWLIRN 551

```

## RESULT 24

AA056702 PRELIMINARY; PRT; 617 AA.  
 ID AA056702  
 AC AA056702  
 DT 02-MAR-2004 (Tremblrel. 27, Created)  
 DT 02-MAR-2004 (Tremblrel. 27, Last sequence update)  
 DT 02-MAR-2004 (Tremblrel. 27, Last annotation update)  
 DE UDP-GalNAc:polypeptide N-acetylglactosaminyltransferase (EC 2.4.1.41).  
 GN PGANTS.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Peerygota;  
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OC NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 EX PubMed12829714;  
 RA Ten Hagen K.G., Tran D.T., Gerken T.A., Stein D.S., Zhang Z.;  
 RT "Functional Characterization and Expression Analysis of Members of the  
 RT UDP-GalNAc:Polypeptide N-Acetylglactosaminyltransferase Family from  
 RT Drosophila melanogaster";  
 RL J. Biol. Chem. 278.35039-35048(2003).  
 DR EMBL; AY268066; AA056702.1--  
 KW Glycosyltransferase; Transferase.  
 SQ SEQUENCE 617 AA; 70482 MW; CA0A7A7A53B73E3 CRC64;  
 Query Match 33.9%; Score 1111.5; DB 2; Length 617;  
 Best Local Similarity 41.9%; Pred. No. 3e-81;  
 Matches 243; Conservative 82; Mismatches 194; Indels 61; Gaps 18;  
 QY 33 ERPDGTGGGSAAVAPAGGSGSHRQKTFPLGDGQK-----LKDMDKEAIRRDAQ 85  
 DB 69 EKSLDGGSG-----ALIMGQ-----FASGISMTPSYVLKKMFLAPSVQEKAG 114  
 QY 86 RVNGGEQR--YPMTDARQDAVRENGFNITYSDKLSLRSLPDIDHPCNSRLEYET 143  
 DB 115 K--GEGMKPKVIRPMDMDLKKKEKQFNLLASDMSLRSRLTDVHRCRRGVASK 172  
 QY 144 LPNTSIIPFNEGSSLLRTVHSLVNSRPELVAYLYVDSPREHLKKPLEDYMLF 203  
 DB 173 LPPTSIIVFNEAWTILRLTWSVYNSRPLALKEIILLVDASERDPLGQLEEVAKL 232  
 QY 204 P-SVRLTKRKEGLIRTRMLGASVATGVTFLDSSHCAVNMVLPPLDRIARRKTI 262  
 DB 233 PVKTFVLRTERKSGIIRARLLGAHNVSEVITFLDAHCEGMLPEPLAIVQRRIV 292  
 QY 263 CPMDIVDHDQFRJETQGDANRGAFFDWMYTKRIPIPE---LQADPSPFESPVMAG 319  
 DB 293 CPIIDVISDEFEYIT-ASDSTWGGFNMKLNFRWYRVPSREMARNNDRTPLRTPWAG 351  
 QY 320 GLPFAVRKMFELGQYDGLGLEWGEQYEISPKVMCGARMEDIPCSRVGHIYRYKVPYK 379  
 DB 352 GLFSDIKQFYEIGSYDEGMIDGEMLEMSFRIOCGGILEIIPCSVGHVFRDKSPYT 411  
 QY 380 VPAGVS--LADNLKRVAVNMDEVAEYIYQRRPEYRLSAGDVAVQKLRSLNCKSPFK 437  
 DB 412 FPGGVAKIVLHNAARVAEVMLEDEWDFYMSGTARKASAGDVSRKALRDLCKKSPRW 471  
 QY 438 FMTKIAMLPRFYRVEPPAAM--GEIRANGTGICADTKIGALGSLRLLEGCVRGREBA 495  
 DB 472 YLENV-----YPSLMLPLDYLYIGIRNAETETCLDTPMKRKNKKYIC--HGGL-- 521  
 QY 496 AMNNQVFTFTWRDIRPGDPQHTKKPCFDALISHTSPVLVLDCHSMKNGQMLKY--RKQXT 554  
 DB 522 ---GNQVAYTKRQDINSDD-----LCLDASSSGPVMNRCHMNGANGQWYDAEKKW 572  
 QY 555 LYHPVSGSCMDCSDHRIFMNT--CNPSSL--TOOMLFE 590  
 DB 573 IRTNMSGCQLRATRD---ANTPLRLPCSYGKGQWLM 609

GL13 MOUSE STANDARD; PRT; 556 AA.  
 ID GL13 MOUSE  
 AC O8CF93; GBLE4; O8BET3;  
 DT 01-OCT-2004 (Rel. 45, Created)  
 DT 01-OCT-2004 (Rel. 45, Last sequence update)  
 DT 01-OCT-2004 (Rel. 45, Last annotation update)  
 DE Polypeptide N-acetylglactosaminyltransferase 13 (EC 2.4.1.41)  
 DE (Protein-UDP acetylglactosaminyltransferase 13) (UDP-  
 DE GalNAc:polypeptide N-acetylglactosaminyltransferase 13) (Polypeptide  
 DE GalNAc transferase 13) (GalNAc-T13) (pp-Gantase 13).  
 GN Name-Galnt13.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OC NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A. (ISOFORM 1), FUNCTION, AND TISSUE SPECIFICITY.  
 RC TISSUE=Brain;  
 RX MEDLINE=22393469; PubMed=12407114; DOI=10.1074/jbc.M203094200;  
 RA Zhang Y., Iwasaki H., Wang H., Kudo T., Kalta T.B., Hennet T.,  
 RA Kubota T., Cheng L., Inaba N., Gotoh M., Togayachi A., Guo J.-M.,  
 RA Hiatsomi H., Nakajima K., Nishihara S., Nakamura M., Marth J.D.,  
 RA Naritoku H.;  
 RT "Cloning and characterization of a new human UDP-N-acetyl-alpha-D-  
 RT galactosamine:polypeptide N-acetylglactosaminyltransferase,  
 RT designated pp-GalNAc-T13, that is specifically expressed in neurons  
 RT and synthesizes GalNAc alpha-serine/chreonine antigen.";  
 RL J. Biol. Chem. 278:573-584(2003).  
 RN [2]  
 RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).  
 RC STRAIN=C57BL/6J; TISSUE=Corpora quadrigemina, and Hypothalamus;  
 RX MEDLINE=22354683; PubMed=12468851; DOI=10.1038/nature01266;  
 RA Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,  
 RA Nishida I., Oosato N., Saito R., Suzuki H., Yamanaka I., Kiyosawa H.,  
 RA Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schonbach C., Gotohori T.,  
 RA Baldarelli R., Hill D.B., Bult C., Hume D.A., Quackenbush J.,  
 RA Schriml L.M., Kanapin A., Matsuda H., Batalov S., Betsel K.W.,  
 RA Blake J.A., Bradt D., Brusic V., Chochia C., Coriani L.B., Cousins S.,  
 RA Dalla B., Dragani T.A., Fletcher C.F., Forrest A., Fraser K.S.,  
 RA Gaasterland T., Gariboldi M., Giesl C., Gostak A., Gough J.,  
 RA Grimmond S., Gustincich S., Hirokawa N., Jackson I.J., Jarvis E.D.,  
 RA Kanai A., Kawai H., Kawasawa Y., Kedzierski R.M., King B.L.,  
 RA Konagaya A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A.,  
 RA Maglott D.R., Maltais L., Marchionni L., McKenzie L., Miki H.,  
 RA Naashima T., Numa K., Okido T., Pavan W.J., Petrea G., Pesole G.,  
 RA Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramachandran S.,  
 RA Ravasi T., Reed J.C., Reed D.J., Reid J., Ring B.Z., Rindgen M.,  
 RA Sadelin A., Schneider C., Semple C.A., Setou M., Shimada K.,  
 RA Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita M.,  
 RA Verardo R., Wagner L., Walestedt C., Wang Y., Watanabe Y., Wells C.,  
 RA Wilming L.G., Wyshaw-Boris A., Yanagisawa M., Yang I., Yang L.,  
 RA Yuan Z., Zavolan M., Zhu Y., Zimmer A., Carninci P., Hayatsu N.,  
 RA Hironaka-Kishikawa T., Kono H., Nakamura M., Sakazume N., Sato K.,  
 RA Shiraki T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukuda S.,  
 RA Hara A., Hashizume W., Imocani K., Ishii Y., Itoh M., Kagawa I.,  
 RA Miyazaki A., Sakai K., Sasaki D., Shibata K., Shinagawa A.,  
 RA Yaunzeshi A., Yoshino M., Waterston R., Lander E.S., Rogers J.,  
 RA Birney E., Hayashizaki Y.;  
 RT "Analysis of the mouse transcriptome based on functional annotation of  
 RT 60,770 full-length cDNAs.";  
 RL Nature 420:563-573(2002).  
 RN [3]  
 RP FUNCTION.  
 RX PubMed=8618846;  
 RA Hennet T., Hagen F.K., Tabak L.A., Marth J.D.;  
 RT "T-cell-specific deletion of a polypeptide N-acetylglactosaminyl-  
 RT transferase gene by site-directed recombination.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 92:12070-12074(1995).  
 RN [4]  
 RP TISSUE SPECIFICITY.  
 RX PubMed=12651884; DOI=10.1093/glycob/cwg062;  
 RA Young W.W., Jr., Holcomb D.R., Ten Hagen K.G., Tabak L.A.;  
 RT "Expression of UDP-GalNAc:polypeptide N-  
 RT acetylglactosaminyltransferase isoforms in murine tissues determined

RESULT 25  
 GL13\_MOUSE



**This Page Blank (uspto)**